

**Ques:el/Orbit**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 19:38:00 ; Search time 1395.77 Seconds  
(without alignments)  
3745.598 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197  
Sequence: 1 ATGATGACATATCGAAGAAC.....ACTACATCATGTCGAATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p11:\*  
10: gb\_p12:\*  
11: gb\_p13:\*  
12: gb\_ro:\*  
13: gb\_sy:\*  
14: gb\_un:\*  
15: em\_fun:\*  
16: em\_hum1:\*  
17: em\_hum2:\*  
18: em\_ln:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: gb\_htg1:\*  
31: gb\_htg2:\*  
32: gb\_in1:\*  
33: gb\_in2:\*  
34: em\_ba1:\*  
35: em\_ba2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: gb\_p14:\*  
39: gb\_hc93:\*  
40: gb\_hc94:\*  
41: gb\_hc95:\*  
42: gb\_hc96:\*  
43: gb\_hc97:\*

44: em\_htg1:\*  
45: em\_htg2:\*  
46: em\_htg3:\*  
47: em\_hum5:\*  
48: gb\_p13:\*  
49: gb\_p15:\*  
50: gb\_htg8:\*  
51: gb\_htg9:\*  
52: gb\_htg10:\*  
53: gb\_htg11:\*  
54: gb\_htg12:\*  
55: gb\_htg13:\*  
56: gb\_htg14:\*  
57: gb\_in3:\*  
58: gb\_htg15:\*  
59: gb\_htg16:\*  
60: gb\_htg17:\*  
61: em\_htg4:\*  
62: em\_htg5:\*  
63: em\_htg6:\*  
64: em\_htg7:\*  
65: em\_hum6:\*  
66: gb\_htg18:\*  
67: gb\_htg19:\*  
68: gb\_htg20:\*  
69: gb\_htg21:\*  
70: gb\_htg22:\*  
71: gb\_htg23:\*  
72: gb\_v11:\*  
73: gb\_v12:\*  
74: gb\_ba3:\*  
75: em\_htg8:\*  
76: em\_htg9:\*  
77: em\_htg10:\*  
78: em\_htg11:\*  
79: em\_htg12:\*  
80: em\_htg13:\*  
81: em\_htg14:\*  
82: em\_htg15:\*  
83: em\_htg16:\*  
84: em\_htg17:\*  
85: em\_htg18:\*  
86: em\_htg19:\*  
87: em\_htg20:\*  
88: em\_htg21:\*  
89: em\_htg22:\*  
90: em\_htg23:\*  
91: gb\_p16:\*  
92: gb\_p17:\*  
93: gb\_sts1:\*  
94: gb\_sts2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query length	DB ID	Description
1	1197	100.0	1197	11 AF069740	AF069740 Homo sapi
2	108.6	9.1	5932	2 BACMENAOUI	M74521 Bacillus su
3	108.6	9.1	5946	2 BACMENAOOP	M74538 Bacillus su
4	108.6	9.1	215640	2 BSUB0016	299119 Bacillus su
5	108.6	9.1	220060	1 AF008220	AF008220 Bacillus
6	103.4	8.6	4662	2 PAX82644	X82644 P.aeruginos
7	89.8	7.5	143051	2 D90911	D90911 Synechocyst
8	76	6.3	1659	2 ECOENTC	M24142 E. coli iso
9	76	6.3	12003	1 AE000165	AE000165 Escherich
10	76	6.3	136742	2 ECUB82598	U82598 Escherichla
11	74	6.2	14027	1 AE004272	AE004272 Vibrio ch
12	72.8	6.1	1302	2 ECOENTCE	M36700 E. coli ito

13	60.8	5.1	1560	2	ECU54790	U54790	Escherichia
14	59.8	5.0	1273	7	AF078080	AF078080	Arabidops
15	57	4.8	2078	8	CRO6065	AJ006065	Catharant
16	51	4.3	1358	2	ECMENGN	U50849	E.coli menf
17	51	4.3	10409	1	AEO00316	AEO00316	Escherich
18	51	4.3	13348	2	D90858	D90858	E.coli geno
19	51	4.3	15676	2	D90857	D90857	E.coli geno
20	49.8	4.2	5166	2	BSU26444	D5444	Bacillus su
21	49.8	4.2	217420	2	BSUB0014	Z99120	Bacillus su
22	48	4.0	4680	2	BACRF01A	M34053	B.subtilis
23	48	4.0	180136	1	BAC180K	D26185	B. subtilis
24	48	4.0	213080	2	BSUB0001	Z99104	Bacillus su
25	47.8	4.0	110000	50	AC020850_1	Continuation (2	Continuation (2
26	46.6	3.9	22070	2	MFCY7D11	Z95120	Mycobacteri
27	45	3.8	164921	7	AF0742186	AF0742186	Cyanidium
28	45	3.8	332375	60	AF0074221	AF0074221	Mus musc
29	44.8	3.7	5057	2	PFBSCEAB	Y03356	Pseudomonas
30	44.4	3.7	8656	2	LACTROP	M87483	L. lactis t
31	44	3.7	12048	1	AEO04162	AEO04162	Vibrio cho
32	44	3.7	12200	74	VCU52150	U52150	Vibrio cho
33	43.6	3.6	2591	2	L81122	L81122	Buchnera ap
34	43.4	3.6	2982	2	SSOSSTRPBG	M98048	Sulfolobus
35	43.4	3.6	6380	2	SSRPRPBG	Z50014	S. solfatar
36	42.4	3.5	351	74	YEN288281	AJ288281	Yersinia
37	41.4	3.5	293350	2	CNSPAY02	AJ288281	Yersinia
38	41	3.4	351	74	YEN270408	AJ270408	Pyrococcu
39	41	3.4	351	74	YEN288282	AJ270409	Yersinia
40	41	3.4	351	74	YEN288284	AJ288282	Yersinia
41	41	3.4	351	74	YEN288285	AJ288284	Yersinia
42	41	3.4	1563	74	STRPPE	AJ288285	Yersinia
43	41	3.4	1563	74	STRPPE	Y01378	Salmonella
44	40.6	3.4	5621	32	AF030155	U01812	S. typhimur
45	40.6	3.4	13548	41	AC015023	AF030155	Drosophila
						AC015023	Drosophila

## ALIGNMENTS

[illegible]

RESULT	1			
AF069740				
LOCUS	AF069740	1197 bp	mRNA	PRI
DEFINITION	Homo sapiens neutral sphingomyelinase mRNA, complete cds.			21-DEC-1999
ACCESSION	AF069740			
VERSION	AF069740.1	GI:660679		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1197)			
AUTHORS	Characterjee,S., Han,H., Rollins,S. and Cleveland,T.			
TITLE	Molecular cloning, characterization, and expression of a novel human neutral sphingomyelinase			
JOURNAL	J. Biol. Chem. 274 (52), 37407-37412 (1999)			
MEDLINE	20069736			
REFERENCE	2 (bases 1 to 1197)			
AUTHORS	Characterjee,S., Han,H., Rollins,S. and Cleveland,T.			
TITLE	Direct Submission			
JOURNAL	Submitted (03-JUN-1998) Pediatrics, Johns Hopkins University School of Medicine, 600 N. Wolfe Street, CMSC 6-124, Baltimore, MD 21287-3654, USA			
FEATURES	Location/Qualifiers			
source	1..1197			

```
CDS
1..1197
1..cirs06_type= kidney
/codon_start=1
/product="neutral sphingomyelinase"
/protein_id="AAFI9052.1"
/db_xref="gi:6606080"
/translat="EMMYHETRALNQSDLIQQLVYALETTTEFGAYFAFPADDTLFFGIG
AIAIKTKQAALOGAVFEGCAVEFSDEQYPOSILMGFWFPEVMTIADKLTFFESDTR
```

QY	781	CTAAGGTCGCCGCCATGCCAAGTTTACTCAAAAAATAAGCAAGTTCACATTTGTACACA	840
Db	781	CTAAGGTCGCCGCCATGCCAAGTTTACTCAAAAAATAAGCAAGTTCACATTTGTACACA	840
QY	841	CCAATTACAGGGGACATTGCGGACATTTAAGTGTGACCCGATGTTGTGACCGCTTCAT	900
Db	841	CCAATTACAGGGGACATTGCGGACATTTAAGTGTGACCCGATGTTGTGACCGCTTCAT	900



OY	901	CCACACGACGACTCGGTGGCTTTCAGAGTGCTCCTATTTGCGATTAATAGT	1020
OY	901	CCAAACACCAGCACTCGGTGGCTTTCAGAGTGCTCCTATTTGCGATTAATAGT	1020
OY	961	GAGAAGACACCTCGGTGGCTTTCAGAGTGCTCCTATTTGCGATTAATAGT	1020
Db	961	GAGAAGACACCTCGGTGGCTTTCAGAGTGCTCCTATTTGCGATTAATAGT	1020
OY	1081	TATATTGCTGGTGGCGGATTTGCCTACATCCCATGCGCAACAAGATATGAAGAACT	1140
OY	1081	TATATTGCTGGTGGCGGATTTGCCTACATCCCATGCGCAACAAGATATGAAGAACT	1140
Db	1081	TATATTGCTGGTGGCGGATTTGCCTACATCCCATGCGCAACAAGATATGAAGAACT	1140
OY	1141	GGGTTGAATTGTGGTGGCATCCGTTCCATGTATGTGAATCAAACGACGAGCAACT	1080
OY	1141	GGGTTGAATTGTGGTGGCATCCGTTCCATGTATGTGAATCAAACGACGAGCAACT	1080
Db	1141	GGGTTGAATTGTGGTGGCATCCGTTCCATGTATGTGAATCAAACGACGAGCAACT	1080
RESULT	2		
BACMENAUI			
LOCUS	BACMENAUI	5932 bp	DNA
DEFINITION	BACMENAUI	menaglutone operon, complete cds.	12-OCT-1994
ACCESSION	M74521		
VERSION	M74521.1	GI:557486	
KEYWORDS	menaglutone.		
SOURCE	Bacillus subtilis DNA.		
ORGANISM	Bacillus subtilis		
REFERENCE	Bacteria: Firmicutes; Bacillus/Clostridium group;		
AUTHORS	Bacillus/Staphylococcus group; Bacillus.		
TITLE	I (bases 1 to 4000); 4001 to 5932)		
JOURNAL	Driscoll,J.R. and Taber,H.W.		
MEDLINE	Sequence organization and regulation of the Bacillus subtilis menB		
FEATURES	Operon		
source	J. Bacteriol. 174 (15), 5063-5071 (1992)		
	Location/Qualifiers		
	1..5932	/organism="Bacillus subtilis"	
	/db_xref="taxon:1423"		
	/germline		
	13..27	/note="putative"	
	/function="Possible regulatory site for pH regulation of promoter"		
-35_signal	13..18		
-10_signal	35..40		
mRNA	47..>579		
	/function-"Message containing ORF1 to possibly ORP5"		
	/evidence-experimental		
RBS	128..133		
	/note="putative"		
gene	142..579		
CDS	/gene="ORF1"		
	142..579		
	/gene="ORF1"		
	/standard_name="ORF1"		
	/note="putative"		
	/codon_start=1		
	/transl_table=1		
	/function="Unknown"		
	/product="Unknown"		
	/protein_id="AA50396.1"		
	/db_xref="GI:557487"		
	/translation="MYTVVORTFRKKVLHLAKAEVNHAVALISYSHQSLDPLSPF		
	NYGKKTGNRFEWSDSELITIGLKEAVFOFNKNSRYREVFQWRPFKFTAFH		
	IYEKKIOMDLICYSDEPLTLFLAKREVINGTISRGISLCRLC"		
gene	582..1547		
	/gene="ORF2"		
CDS	582..1547		

```

/gene="ORF2"
/standard_name="ORF2"
/note="putative"
/codon_start=1
/transl_table=11
/function="unknown"
/product="unknown"
/protein_id="AA50397.1"
/db_xref="GI:357488"
/translation="MTAGPEPLTVNRWVSGEDA EAVLEGLKAFAEFVNDPEQEDQ
AVIAAEELDKDMLKAIETATSOIKEQYKVLARELLITPGPIQIPVLKTLTLD
DOGTVEALIEEGTKEFVGASPERLIKEDGTWVSLGASIKGVNEEDRRIGLET
INDLEKILHDIYGVGHIMINAFVSSCSPEKEDGVLTKTTSVGHLPPIYGOIAPESAS
IPDLKELHPPPALDGSPQEKAVDYIRIEPMNSGVAALPTGIRIDSDNCEFAVALRS
GLIEGTALKEPFGGCIYEDSEPISEYETQIKLPMIASIGERR"
1624..2049
/gene="memp"
1624..2049
/gene="memp"
/standard_name="2succinyl-6hydroxy-2,
4cyclohexadiene-1carboxylate synthase"
/codon_start=1
/transl_table=11
/function="Convert Isochorismate to SHCHC"
/evidence=experimental
/protein_id="AA50398.1"
/db_xref="GI:557489"
/translation="MSGKLYTLAVLCNAHPDISVHVQIDERSAGFPALGLAKAKORP
VLLICTSTAAANFPYPAVEAHYSRPIIVLTADRPHELRVGAAPALINQHLFENFV
KEFTDSALPEESPQMLRIIRTLASRAGEAQRKRMALSM"
2072..3187
/gene="mencf"
2072..3187
/gene="mencf"
/standard_name="mencf"
/codon_start=1
/transl_table=11
/evidence=experimental
/protein_id="AA50399.1"
/db_xref="GI:557490"
/translation="MPDISDEPFGMRGHRVSVKTCOSVDRESISDAVEMLAFAEK
GMIVGELHSDAKENITATLSKALOYPLADPLSNLNGCHDKSTYIDAYDSLKDE
LKRKRIPDVILKFGPMVPKSPVFLMKDDPTIQIVIDEGGWDRPTQASAHMHCNA
SVFAEIMAGLTAATRSSEMLEKQFVNGRREHLQTISSBDSFEGNLRLIQLHLVP
ENSSFLVNSNDIIRVDYTFPEKORRPRITYNRRGANGIDGVSSAMGVCGCTKAPVTL
VYGDLSFHDNLGLIATAKKIGLPIVLIVNNDGGGIFSLPQASEKTHFEDLGTPTG
IDFKMAALYGGTYSCPASNDPEFKTAAKROTSPDSI"
3270..3857
/gene="ORF5"
3270..3857
/gene="ORF5"
/standard_name="ORF5"
/note="putative"
/codon_start=1
/transl_table=11
/function="unknown"
/product="unknown"
/protein_id="AA50400.1"
/db_xref="GI:557491"
/translation="WGTVNITVSDGVRVYVADECPNASEAVYCLHGTGTSKQSWTFED
EMLPSRLIKIDCLGHTDAPLNGKRKYSTTRVSDLAELFDQLKLRVLIQYSMGQ
RLATSEFAMTYPSGYRHCLICAKARRSGSKHNGNESCGETSLILTFEYETSKRLIRIG
RSLSEFHSRGLKIFGQVBRVAVCGTIRGLRTV"
4089..4103
/note="putative"
/function="Sequence may be involved in pH regulation of
promoter."
/phenotype="Unknown"
4102..4107
4125..4130
4138..55722
/gene="menb"
/evidence=experimental

```

BASE COUNT	1594	a	1248	c	1552	g	1538	t
gene	4138 . :5722	/gene="menB"	4148 . :5722	/gene="menB"	4148 . :5722	/gene="menB"	4148 . :5722	/gene="menB"
RBS	4148 . :5722	/gene="menB"	4148 . :5722	/gene="menB"	4148 . :5722	/gene="menB"	4148 . :5722	/gene="menB"
CDS	4157 . :4942	/note="putative"	4157 . :4942	/note="putative"	4157 . :4942	/note="putative"	4157 . :4942	/note="putative"
	4157 . :4942	/gene="menB"	4157 . :4942	/gene="menB"	4157 . :4942	/gene="menB"	4157 . :4942	/gene="menB"
	4157 . :4942	/standard_name="Men B"	4157 . :4942	/standard_name="Men B"	4157 . :4942	/standard_name="Men B"	4157 . :4942	/standard_name="Men B"
	4157 . :4942	/codon_start=1	4157 . :4942	/codon_start=1	4157 . :4942	/codon_start=1	4157 . :4942	/codon_start=1
	4157 . :4942	/transl_table=1	4157 . :4942	/transl_table=1	4157 . :4942	/transl_table=1	4157 . :4942	/transl_table=1
	4157 . :4942	/function="Convert OSB-CoA to DNA in Menaquinone biosynthesis"	4157 . :4942	/function="Convert OSB-CoA to DNA in Menaquinone biosynthesis"	4157 . :4942	/function="Convert OSB-CoA to DNA in Menaquinone biosynthesis"	4157 . :4942	/function="Convert OSB-CoA to DNA in Menaquinone biosynthesis"
	4157 . :4942	/evidence=experimental	4157 . :4942	/evidence=experimental	4157 . :4942	/evidence=experimental	4157 . :4942	/evidence=experimental
	4157 . :4942	/product="dihydroxyapatitic acid (DHNA) synthetase"	4157 . :4942	/product="dihydroxyapatitic acid (DHNA) synthetase"	4157 . :4942	/product="dihydroxyapatitic acid (DHNA) synthetase"	4157 . :4942	/product="dihydroxyapatitic acid (DHNA) synthetase"
	4157 . :4942	/protein_id="AA50401.1"	4157 . :4942	/protein_id="AA50401.1"	4157 . :4942	/protein_id="AA50401.1"	4157 . :4942	/protein_id="AA50401.1"
	4157 . :4942	/db_xref="GI:557492"	4157 . :4942	/db_xref="GI:557492"	4157 . :4942	/db_xref="GI:557492"	4157 . :4942	/db_xref="GI:557492"
	4157 . :4942	/translation="MKIFETDNYGIAKTTINRPNVHNAFTPKVAEMIDFADARDDO NGVATVLGAGADKAFCSGDQKVRGSGYVGDQIPRLNVLQRLIRVIRPKPVAMAM SGVIGGGVHLITICDLTIADNADIPGOTGKVGSPAGVSGYLARIVGKKRREIM YICGVNAOEALDMGLVNTVPLLEOLEERTKCEEMLEKSPALRLIKAFFNADTDG LAGIQQFAGDNTLTYTTDEAKESGRDSFKERKRPFCQFPFP"	4157 . :4942	/translation="MKIFETDNYGIAKTTINRPNVHNAFTPKVAEMIDFADARDDO NGVATVLGAGADKAFCSGDQKVRGSGYVGDQIPRLNVLQRLIRVIRPKPVAMAM SGVIGGGVHLITICDLTIADNADIPGOTGKVGSPAGVSGYLARIVGKKRREIM YICGVNAOEALDMGLVNTVPLLEOLEERTKCEEMLEKSPALRLIKAFFNADTDG LAGIQQFAGDNTLTYTTDEAKESGRDSFKERKRPFCQFPFP"	4157 . :4942	/translation="MKIFETDNYGIAKTTINRPNVHNAFTPKVAEMIDFADARDDO NGVATVLGAGADKAFCSGDQKVRGSGYVGDQIPRLNVLQRLIRVIRPKPVAMAM SGVIGGGVHLITICDLTIADNADIPGOTGKVGSPAGVSGYLARIVGKKRREIM YICGVNAOEALDMGLVNTVPLLEOLEERTKCEEMLEKSPALRLIKAFFNADTDG LAGIQQFAGDNTLTYTTDEAKESGRDSFKERKRPFCQFPFP"	4157 . :4942	/translation="MKIFETDNYGIAKTTINRPNVHNAFTPKVAEMIDFADARDDO NGVATVLGAGADKAFCSGDQKVRGSGYVGDQIPRLNVLQRLIRVIRPKPVAMAM SGVIGGGVHLITICDLTIADNADIPGOTGKVGSPAGVSGYLARIVGKKRREIM YICGVNAOEALDMGLVNTVPLLEOLEERTKCEEMLEKSPALRLIKAFFNADTDG LAGIQQFAGDNTLTYTTDEAKESGRDSFKERKRPFCQFPFP"
terminator	4963 . :4993	/gene="menB"	4963 . :4993	/gene="menB"	4963 . :4993	/gene="menB"	4963 . :4993	/gene="menB"
	4963 . :4993	/evidence=experimental	4963 . :4993	/evidence=experimental	4963 . :4993	/evidence=experimental	4963 . :4993	/evidence=experimental
misc_feature	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"
	4975 . :4982	/standard_name="Golden Loop' sequence"	4975 . :4982	/standard_name="Golden Loop' sequence"	4975 . :4982	/standard_name="Golden Loop' sequence"	4975 . :4982	/standard_name="Golden Loop' sequence"
	4975 . :4982	/function="possible stabilized of RNA secondary structure."	4975 . :4982	/function="possible stabilized of RNA secondary structure."	4975 . :4982	/function="possible stabilized of RNA secondary structure."	4975 . :4982	/function="possible stabilized of RNA secondary structure."
	4975 . :4982	/evidence=experimental	4975 . :4982	/evidence=experimental	4975 . :4982	/evidence=experimental	4975 . :4982	/evidence=experimental
	4975 . :4982	/product="None"	4975 . :4982	/product="None"	4975 . :4982	/product="None"	4975 . :4982	/product="None"
	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"
	4975 . :4982	/note="putative"	4975 . :4982	/note="putative"	4975 . :4982	/note="putative"	4975 . :4982	/note="putative"
	4975 . :4982	/function="possible recognition sequence for pH regulation of promoter."	4975 . :4982	/function="possible recognition sequence for pH regulation of promoter."	4975 . :4982	/function="possible recognition sequence for pH regulation of promoter."	4975 . :4982	/function="possible recognition sequence for pH regulation of promoter."
	4975 . :4982	/phenotype="Unknown"	4975 . :4982	/phenotype="Unknown"	4975 . :4982	/phenotype="Unknown"	4975 . :4982	/phenotype="Unknown"
	4975 . :4982	5002 . :5007	4975 . :4982	5002 . :5007	4975 . :4982	5002 . :5007	4975 . :4982	5002 . :5007
	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"
	4975 . :4982	5024 . :5029	4975 . :4982	5024 . :5029	4975 . :4982	5024 . :5029	4975 . :4982	5024 . :5029
	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"
	4975 . :4982	5038 . :5932	4975 . :4982	5038 . :5932	4975 . :4982	5038 . :5932	4975 . :4982	5038 . :5932
	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"	4975 . :4982	/gene="menB"		

OY	521	TAGCTGACGAGCGCAATACGTATCATGTGGTTTAAACGCACATGATGATATTGTTATT	580
Db	883	TGGAGATCAGCAGCAACGATATGTTTTTGCAATTTGAACAAGAAGCAAAACCTTTTGTGC	942
OY	581	CAGCAACACCGAACCAGTGTAGTGGCTATGTGACGGTGTCAATCGCTACCGCGCGGTGC	640
Db	943	GCGCTCTCCGGAAGACTGATCAAABAAGACGGCGCACATGTATCTTCCTCTGTCTGG	1002
OY	641	CTGGGACAAGCCGGCGGGACGATGGCGCTGACGATATCGCGTTTAGCGCAAGCGTTGT	700
Db	1003	CAGGCTCCATTAAACGAGAGCGGTGATGAAGAAGACGACGCCGCGATAGCCCTTGATTAAT	1062
OY	701	TAGCCAGTCAGAAAAACCGCATTTGAACATCATATGTGTGGCAAGTATCAAGACACCT	760
Db	1063	TAAACGATGAGAAAAAACCTGCTGTGACATGATATTGTGTGTATGATATATAATGCTT	1122
OY	761	TGCAGAGCTGACGACGCTGCGCTAAAGTGTCCGGCCATCCCAAGTTTACTCAAAAATAAGC	820
Db	1123	TTGTATCAAGCTGTTTCAAGGTTGAAAAGCCTGATGAGCTGTTTGTACAAAACAAA	1182
OY	821	AAGTTCACGATTTTGTATCACACCAATTTACAGGGGACATTCGCGCACATTTAAGTGTGACG	880
Db	1183	GCGTTCACACACTGTTTATCGCGCATTTGTGCGACAGCTCGGGAGTCTGCTTGATTTTGG	1242
OY	881	CGATTGTGTACCGCTTGTCATCCACACACGACGCTGGTGGCGTCCCACGTAGACGGCCCC	940
Db	1243	ACTTATTTAGAAAGCTGCACCCCACTCCCGGCTTTGGGGATCATCCTGAGAAAAAGCTG	1302
OY	941	TGTATTACATTTGCCGACCATGGAAGACACGTCGTGGGCTTGTTCAGGTGCTTAATGGCT	1000
Db	1303	TTGACGTGATCAGGAGATTTGACCGATGTCCCGGGCTGTATGCGGCCCTTATCGGCT	1362
OY	1001	ATTTTACCGCAGATPATATGTAGGGGAATTTGTGTGGCATCCGTTCCATGTATGTGAATC	1060
Db	1363	GGATTGACAGCAGAGATPAACGGGGAATTCGGGGTGCACATTGTTCAAGGGCTCATCGAAG	1422
OY	1061	AAACGACACGACGACCACTTATTTCGTGTGCGGGGATTTGTGCTATCTCCATCGATGCGC	1120
Db	1423	GAA-----GCAACAGCAAGGCTGTTTGGCGGGGTGGGATTTGTGGAMACATCGGAACCA	1476
OY	1121	AACAGATPATGTGAAGAACTGGTGTGAATTTGAACCCATG	1161
Db	1477	TCTCGGATATGTAGAAACACACAGATTAATTAGMAGCCGATG	1517
RESULT	3		
BACMENAOOP			
LOCUS	BACMENAOOP	5946 bp	DNA
DEFINITION	Bacillus subtilis menaquone operon: menF, menD, menB and menE genes, complete cds.		
ACCESSION	M74538	M74182	M74183
VERSION	M74538.1	GI:1185287	
KEYWORDS	menaquone.		
SOURCE	Bacillus subtilis (strain RB1)		DNA.
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus.		
REFERENCE	1 (bases 1 to 5946)		
AUTHORS	Miller,P., Mueller,J., Hill,R. and Taber,H.		
TITLE	Transcriptional regulation of a promoter in the men gene cluster of Bacillus subtilis		
JOURNAL	J. Bacteriol.	170 (6),	2742-2748 (1988)
MEDLINE	88227858		
REFERENCE	2 (bases 1 to 5946)		
AUTHORS	Driscoll,J.R. and Taber,H.W.		
TITLE	Sequence organization and regulation of the Bacillus subtilis menBE operon		
JOURNAL	J. Bacteriol.	174 (15),	5063-5071 (1992)
MEDLINE	92332443		
REFERENCE	3 (sites)		
AUTHORS	Therk,C., Gauss,P., Thernes,C., Grobe,D.R., Gayle,M., Guild,N., Stormo,G., d'Audenton-Carafa,Y., Uhlenbeck,O.C., Tinoco,I. Jr., Brody,E.N. and Gold,L.		

TITLE CUUCCG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (5), 1364-1368 (1988)  
MEDLINE 88144437  
REFERENCE 4 (bases 1 to 3913)  
AUTHORS Rowland, B., Hill, K., Miller, P., Driscoll, J. and Taber, H.  
TITLE Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes  
JOURNAL Gene 167 (1-2), 105-109 (1995)  
MEDLINE 96144257  
REFERENCE 5 (bases 1 to 5946)  
AUTHORS Rowland, B.  
TITLE Direct Submission  
COMMENT Submitted (21-OCT-1991) Belinda M. Rowland, Division of Infectious Disease, Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201-2002, USA  
FEATURES On Feb 15, 1996 this sequence version replaced gi:143180.  
source  
1. .5946  
/organism="Bacillus subtilis"  
/strain="RBJ"  
/db\_xref="taxon:1423"  
14. .28  
/note="putative"  
/citation=[1]  
/function="regulatory site for pH regulation of promoter"  
14. .19  
36. .41  
48. .>590  
/note="5' RNA terminus mapped by high-resolution S1 nuclease protection assays (J. Bacteriol. 170, 2742-2748, 1988); exact 3' terminus unclear, but transcripts into menB are undetectable; contains ORF1 to possible ORF5"  
/citation=[1]  
/evidence="experimental"  
130. .134  
143. .1558  
143. .1558  
/gene="menB"  
/note="menB"  
/note="based on similarity to E. coli EntC, A. hydrophila AmoA, B. subtilis Dhbc"  
/codon\_start=1  
/transl\_table=11  
/product="isochorismate synthase"  
/protein\_id="AAC37013.1"  
/db\_xref="GI:1185288"  
/translation="MTVTQRTFRKEVLAHLHAKAEVNHAVLISYSGJESLDPLSF NYGAKKYNRFFMSDPESELITVIGIKFAVOTNKNSERYREVEQMERKKTAFH IYEEKIQHSAVGPVLFSGFSFDPCEKRSQMDHRSQDPFYPALMLTMTAEGPLITV NFWVSGGEAEVAVLBEGLKFAAEFVWPDKEQAVIAAEELDKDMLKALETATSQ IREKQYDKVLAARELLTPDGPLOIEPVAKTLIDDOTSYVAIEDEGTFVGAEPER LIRKDGVTWSSCLAGSIRKGVNEEDDRIGLELDEKNLELHDIVGMINHAEPSS CSEVEKPDGPVLYKTSVOHLFTPIVIGOLRESASIFDLKLPFPALGSGPOEKVVD VRIETEPMSRGVYAAPIGWIDODNGEFVAVALRSGLIEGSTARLFRAGCGIVEDSEIS EYEETQIKRKPMISALGERR"  
1541. .3201  
/gene="menD"  
1541. .1545  
/gene="menD"  
1555. .3201  
/gene="menD"  
/note="based on similarity to E. coli MenD and enzymatic analysis of mutants (J. Bacteriol. 176, 2648-2653, 1994)"  
/codon\_start=1  
/transl\_table=11  
/product="2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase"  
/protein\_id="AAC37014.1"  
/db\_xref="GI:1185289"  
/translation="MTVNPITHTYIGSFIDEFALSGITDAVVCGRSPTPLAVLCANP DISVHOIDERSAGFALGALAKAKORPVLLICTSGTAAANFYPAVEEAHVSPIITVL TADRPHELRAVCAPOAINHPLFGNFVKFTTSALPPESPQMLRTYIRTLPSRAAGAAO

KRPMGVHVNPLREPLMDPLSDPEPRRMRTGHRVSKTGTOGVDSRESLSDYAEMLAE ARKMVIVGCELSHDADKENIITLSTKALOYPIIADP.SNIRNCYCHKSTVYIDAVIDFLK DBELKRLKRPDIYIRFGPMRVSKPVLMLKDDPTIQYITIDEGGCRDPTQASAMHIC CNASVFAEIEIMAGLITATRRSSEWLEKMFVNGFREFHLOTTISEDVSFEGNLYRILOH LVPENSLSLFGNSMPTIRVDYTFEKKODRPFRIYSNKGANGIDGVSSAGVCEGTAP VLVILDELSFEYHINGLINLAARKIGIPLYIVLVNDDGGJFSEFLPQASETHEDELFCT PGVLDFKHAALAYGTYISCPASWDEKFTAVARQRTSPDSI"  
3269. .3273  
3284. .3871  
/gene="ORF3"  
3284. .3871  
/gene="ORF3"  
/note="similar to triacylglycerol lipase of P. immobilis"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC37015.1"  
/db\_xref="GI:143185"  
/translation="MGTVNITVSDGYRAVADEGPMASPAVVCILHGFTGKOSWPFID EMLPDSRLKIDICLGHGERTDAPLNGKRYSTTRVSDPLAEIPDLKHLKVKLYSGMG KLAYSFAMTPYSGYRHLCLKARRRSGKHLGNGNESCGETSLLILEYETGSKRLNIG RLSLCFHLRSGWIKIFGTGYDRAVCGTIRGLRTV"  
4103. .4117  
/note="sequence motif seen in promoter regions for several B. subtilis genes involved in TCA cycle and electron transport: putative"  
/citation=[2]  
/function="involved in pH regulation of promoter"  
/phenotype="unknown"  
4116. .4121  
/gene="menB"  
/note="menB"  
/note="identified by appropriate spacing from mapped 5' menB RNA and homology to B. subtilis vegetative type promoters"  
/citation=[2]  
4116. .5736  
/gene="menB"  
4139. .4144  
/gene="menB"  
/note="menB"  
/note="identified by correct spacing from a mapped 5' end of menB transcripts"  
/citation=[2]  
4152. .>5736  
/gene="menB"  
/note="menB"  
/note="5' RNA terminus mapped by high resolution primer extension assays.; Exact 3' terminus is unclear but extends at least through menE."  
/citation=[2]  
/evidence="experimental"  
4162. .4166  
/gene="menB"  
/note="menB"  
/note="putative"  
/citation=[2]  
4171. .4956  
/gene="menB"  
/note="menB"  
/note="predicted MW of 28584.4 Daltons, 261 a.a. residues, estimated pI= 5.99"  
/citation=[2]  
/codon\_start=1  
/transl\_table=11  
/function="convert OSB-CoA to DHNA in menaquinone biosynthesis"  
/evidence="experimental"  
/product="dihydroxyphenolic acid (DHNA) synthetase"  
/protein\_id="AAC37016.1"  
/db\_xref="GI:143186"  
/translation="MKIEPDTNGIKITINRPEVNAATPTKVAEMINAFDARDQO NNGVILAGAGDAFCSSGDDKVRGKGTVGDDQIPRLNVLDLQRLIRITPVPVAAV SGVAIGGCVHLHIVCDTITADNAIJGQGPVKSGSDAGVSGYLARIYGHKRAEIV YLCROYNAOEALDMGLVNTVVPLEOLEEETIKWCEMLKESPTALRFLKAANADTDC LAGIOOFAGDADTLVYTTDEAKERGDSFKEKKRPDRGQPRPP"  
4977. .5007  
/gene="menB"

terminator



gene  
CDS  
/transl\_table=11  
/protein\_id="CAB14906.1"  
/db\_xref="GI:2635412"  
/translation="MKKLIVAFILLSQLIDGSLGMAVGVTSILLAPGIPPAVANS  
ASVHIAEVYTVASGVSHIKFGNDVKQVYLVLDGSLGALGAAPLSQIPGDVAKPY  
ISLPIILLAGVYLIRFLPYKPALEKRVNKRKOSIPUGVIRADATGGCGRPVTT  
TPILSRKGLSPKRVAVGIVDTSEFAIAVSATAGFLISGSDVWMLVFSIMAGGIIAT  
AIVAMLVKQHPQMLMGVLVGGFIIILVNAKRLINEMIANVAVHPLIYTAIGAIVLSAM  
LEVLKIGRNIVKIKTAVDVHLKEK"  
complement(1020. .2270)  
/gene="hlp0"  
complement(1020. .2270)  
/gene="hlp0"  
/EC\_number="3.5.1.32"  
/codon\_start=1  
/transl\_table=11  
/product="hippurate hydrolase"  
/protein\_id="CAB14907.1"  
/db\_xref="GI:2635413"  
/translation="MSLDYMRNIEGSEYPTGTNDILTLKEESNPVNLSTLEKOLIGG  
RHLHGVPELSKEEPEETAPFKKLEKGIQIRPAKTKGVFADIAGSEBPALALRA  
DIDAPRIEKTGYLPASKKHKGIMHACGDFHTAALLKAPFLIKENOSLSKATYLLQ  
PAEEAGACTYGLDEGOLDGIDVATGLHNKPDIAVGLTKGTPUMAADVETKYLEEG  
KGAAHALPNEGDPPIIGASOLIVALTQYSRNVNPLDSALITVKINGSTWYVIPDT  
VVEESTVTFDESVENQVKORFEFAVTEQISAAFSIKANVKWMSGPPLCNDSEATITGLV  
RAAKKAKLOVIDPAPSTAGDEFAVYLLHPIGSPAFCTGDDHDMHNPATIDETAII  
KASYLYESAKRLDLSNEKISD"  
complement(2288. .2980)  
/gene="fibr"  
complement(2288. .2980)  
/gene="fibr"  
/function="regulation of riboflavin biosynthesis genes"  
/note="alternate gene name: ytnk"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="CAB14908.1"  
/db\_xref="GI:2635414"  
/db\_xref="SWISS-PROT:P94465"  
/translation="MTITAGTVYKQKQLGRKLPPTANVDAKIHGLNGVYGVATVNN  
HOFHLGVNIDGVKPTVGSNLEKLTLEIFLDFHRDLYGKELICSLFKIRERRDLSLE  
STLQKIDKQIPDQAKRPFLIGIMAPNKRKESLSHQELNLPDLCEYKCKNNLYGVNRGV  
YVINDNFEFGSITQVAVRYRIYLSLSFLKEDNPVKVSKYIRFGAGLADKLNRPLIS  
SVSESEENIG"  
complement(3027. .4355)  
/gene="ytnJ"  
complement(3027. .4355)  
/gene="ytnJ"  
/function="unknown"  
/note="similar to nitriIotriacetate monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14909.1"  
/db\_xref="GI:2635415"  
/db\_xref="SPTREMBL:O34974"  
/translation="MTRADPIQFMGIMHGVGTTDGMWRHPVDPSASTNIEFYVKKAO  
TAEKLESPFIPLADGLPISEKSIPIHFLKRPETITLLASVYTNIGLVGTFSTPE  
PFTISLQMSLDHISGGRAGNVLVTSPEQASARNHNSKINLEPHEREKIAOELHADVVR  
GLMNSWEMDAPLHNKRTGQFPDQALHLSLNKRGKGFVQEGELNIGRSKOGSPVVFQ  
SETEGROCAAKNADAIFTHNSNLSLETKAFVADVYSRAADEGSDSSVRIIPGISPIYA  
DTEEAKEKRYREFALDIPINNAVYTLARFPDQDLSVYPLDEPPIIDGDKNMFQST  
TDRIKREKARNLTLREVAOGMAPPRTLFICTPPRVALSLETMPNARAOEIVGSDI  
PCMTLDAPEKVIPIIQEKGTYRQDYRGSTLRENIGLGIPOHOSVLSHSHH"  
complement(4352. .4633)  
/gene="ytnI"  
/gene="ytnI"  
complement(4352. .4633)  
/gene="ytnI"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB14910.1"

```

gene
CDS
/db_xref="GI:2635416"
/db_xref="SPTREMBL:O34639"
/translation="MSDVNIVYWSKKGCSYCEEVKNLYLNKEGFPNIDIVSEKRLR
DLYQKYGVRHYPVVEIGRGNQOGITGEIHLDLALANHAQIKKAER"
complement(4648..5652)
/gene="ytm0"
complement(4648..5652)
/gene="ytm0"
complement(4648..5652)
/gene="ytm0"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14911.1"
/db_xref="GI:2635417"
/db_xref="SPTREMBL:O34846"
/translation="MRLSIIDLSLIGSEPTADTLQHPYKLAQMAEEGCHFPWVE
HNNDEIAGSAPEVILGYLAASRTKRLRSLGCMLOHYSYKAEQPHLLSALAPGI
DLGVRKAPGGFOLDLDAETKRPVPRODERKELELTHFVRDPDPHRTAVALPRQ
VDRKPGFLLTGSTESIAISAKGISFEVAYFLGEEVLEKARAFDAHLPGSENE
FHLPARVFAHRTKEAEKHIVSRESIKVVLKDRKYNVSGEQAAYLEAVTEPYDI
YOKTIIAGTKEEVABELTRLSGTYKINDFVIPTPIKNAVEKOLSTQLSDVLAAR"
complement(5649..6428)
/gene="hisp"
complement(5649..6428)
/gene="hisp"
/codon_start=1
/transl_table=11
/product="histidine transport protein (ATP-binding
protein)"
/protein_id="CAB14912.1"
/db_xref="GI:2635418"
/db_xref="SPTREMBL:O34900"
/translation="MEIKNIHQPCIIHHVILGINTYRKEGVVTIIGPSGKPTFL
KCLNLEPRDEGIIISHDKVINCRFSKSEVHLKROTAWFQOYHLFAKTVIENW
EGULTARKRKQDAVAANAENLEKVELQDLNAYPSQLSGQORVGIATVALIHDPV
LELFDPTAIDELVEVLEMLLEIKTATMTVYVHSEMFARVSDQVVEMGVIV
EGQPEEVRHRTKDRTRQFLRRVSPDYLFEPKEHIKEPVI"
complement(6425..7132)
/gene="ytm0"
complement(6425..7132)
/gene="ytm0"
complement(6425..7132)
/gene="ytm0"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB14913.1"
/db_xref="GI:2635419"
/db_xref="SPTREMBL:O34931"
/translation="MQDFEPFIVASAKEMVKTIPLTMAVLPPIVFGFLVALNITVR
IFRIGIVACSRVYSFPPSTPAIILHMLIYIGIPPAOVSSFFHLGSALEPIYSM
FYVALSLTRAGAYLVEIIRSGILAMDQGEAVYSIGLTSQFRFVILPOALKVPID
NFTNNGIGFLHTLSIAAIVAVPEITGTATIVASDNYAFLFAIGAAIYVWLTLLIES
ANGVLERKARROGGTL"
complement(7162..7881)
/gene="ytmL"
complement(7162..7881)
/gene="ytmL"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
Query Match
Best Local Similarity 9.1%; Score 108.6; DR 2; Length 215640;
Matches 316; Conservative 0; Mismatches 319; Indels 6; Gaps 1;
QY 521 TAGCTGACGACGGCAATACGATCATGTGGTTTAAAGCAGCATGATGATTTGTTATTj.580
DB 154276 TGGAGCATCAGCAGCAACACTATCTTTTTCATTTGACACAGAAGCAAAACCTTTGTC 154277
QY 581 CAGCAACACCGGACGGTATTAGTGGCTATGTCAAGTGGTGTCAGATCGTACGGCGGGGTGC 640
DB 154216 GCGCGCTCTCCGGAAGACATGATCAAAAGAGACGGCGCATCTGTCATCTTCCGTGTGG 154217

```

Oy	641	CTGGACAAAGCCGGCGGCGGCAGTATGGCCTCATGCAATTTCGGCTTAAGCGAAGCCGTTCG	700
Db	154156	CAGGCTCATTAAACGAGGCGCTGATBAGAAGACGCCGCCGAGTAGGCGCTTGAAATTAT	154097
Oy	701	TAGCCAGTCAGAAAAACC CGCATTTCAATCATATATGTCGTGGCAAGTATCACGACGCT	760
Db	154096	TAAACGATGAGAAAAACCTGCTTGATGATGATATTTGTTGGTAGATATGATTCATAATGCTT	154037
Oy	761	TGCAAGACGTACACAGCTCGCTAAAAGTGGCCGCGCATGCCAAGTTTACTCAAAAATAAC	820
Db	154036	TTGTATCAAGCTGTTCAGAGGTTGTA AAAAGCCTGATGACACTGTTTTGTACAAAACA AAAA	153977
Oy	821	AAGTTTCAGCATTTTATACACACCAATTTACAGGGGACATTTGGGGACATTTAAGTGAACG	880
Db	153976	GCGTTTCACGACCTCTTTTACGCCGATTTTCGACAGCTGCGCGAGTCTGCTTCGATTTTGG	153917
Oy	881	CGATTTGTACCGCGCTGTCATCCACACACACACACATCGGTGGCGTCCACAGCTGAAGCGCCC	940
Db	153916	ACTTAATATGACAGCTGCAGACCCACCTCCCGCGGTGGGGATGACCTCAGAGAAAAAGCTG	153857
Oy	941	TGTAATTACATTGGCACCATTGAGAAGACACTGTGGCTTGTTCGACAGTCTTATTTGGCT	1000
Db	153856	TTGACGTGATCAGGAGATTGAACCGATGTCCCGCGCTGTATGGCGCCCTATGCGCT	153797
Oy	1001	ATTTTACCGCAGATTAATAGTGGGAATTTGTGGTGGCATCGCTTCATGATATGTAATC	1060
Db	153796	GGATTGTACAGCCAGAGATTACGGGAAATTCGCGTGGCATTTCTTTACAGGGCTATATGAAG	153737
Oy	1061	AAACGACGCGACGACCACTTATTTGTCGTGTGCCGGGATTTGGCTGACCTCCGATGCGC	1120
Db	153736	GAA-----GCACGACAGAGCGTGTTCGCCGGTGGCGGTATTTGGAAAGACTCGGAACCAA	153683
Oy	1121	AACAAGATATGAGAAACCTGGCTTGAATTTGAACCCAG	1161
Db	153682	TTCTGGAAATATGAGAAACACAGATTAAATTTGAAGCCGATG	153642
RESULT	5		
LOCUS	AF008220		
DEFINITION	Bacillus subtilis rmb-dnaB genomic region.	04-FEB-1998	
VERSION	AF008220		
KEYWORDS	AF008220.1 GI:229135		
SOURCE	.		
ORGANISM	Bacillus subtilis. Bacillus subtilis. Bacteria, Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus. 1 (bases 1 to 7430) Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,K.F. Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon rnb		
REFERENCE	JOURNAL Gene 37 (1-3), 261-266 (1985)		
MEDLINE	86031361		
REFERENCE	2 (bases 153210 to 153762)		
AUTHORS	Connors,M.T., Mason,J.M. and Setlow,P.		
TITLE	Cloning and nucleotide sequencing of genes for three small, acid-soluble proteins from Bacillus subtilis spores		
JOURNAL	J Bacteriol. 166 (2), 417-425 (1986)		
MEDLINE	86195826		
REFERENCE	3 (bases 213161 to 218473)		
AUTHORS	Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.		
TITLE	Nucleotide sequence and organization of dnaB gene and neighborhood genes on the Bacillus subtilis chromosome		
JOURNAL	Nucleic Acids Res. 14 (24), 9989-9999 (1986)		
MEDLINE	87117549		
REFERENCE	4 (bases 200404 to 201481)		
AUTHORS	Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.		
TITLE	Cloning and nucleotide sequence of phop, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis		
JOURNAL	J Bacteriol. 169 (7), 2913-2916 (1987)		
MEDLINE	87250247		

REFERENCE AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE	5 (bases 201234 to 202122) Seli, T., Yoshikawa, H., Takahashi, H. and Saito, H. Nucleotide sequence of the <i>Bacillus subtilis</i> phoK gene J. Bacteriol. 170 (12), 5935-5938 (1988) 89053932 6 (bases 142232 to 144147) Grundy, F.J. and Henkin, T.M. Cloning and analysis of the <i>Bacillus subtilis</i> rpsD gene, encoding ribosomal protein S4 J. Bacteriol. 172 (11), 6372-6379 (1990) 91035248 7 (bases 133624 to 134990) Henkin, T.M., Grundy, F.J., Nicholson, M.L. and Chambliss, G.H. Catalytic repression of alpha-amylase gene expression in <i>Bacillus</i> <i>subtilis</i> involves a trans-acting gene product homologous to the Escherichia coli lacI and galK repressors Mol. Microbiol. 5 (3), 575-584 (1991) 91260441 8 (bases 140810 to 142610) Henkin, T.M., Glass, B.L. and Grundy, F.J. Analysis of the <i>Bacillus subtilis</i> tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes J. Bacteriol. 174 (4), 1299-1306 (1992) 92138624 9 (bases 217570 to 220060) Putzer, H., Gendron, N. and Grunberg-Manago, M. Co-ordinate expression of the two threonyl-tRNA synthetase genes in <i>Bacillus subtilis</i> : control by transcriptional antitermination involving a conserved regulatory sequence EMBO J. 11 (8), 3117-3127 (1992) 92347349 10 (bases 134990 to 141290) Grundy, F.J., Waters, D.A., Takova, T.Y. and Henkin, T.M. Identification of genes involved in utilization of acetate and acetoin in <i>Bacillus subtilis</i> Mol. Microbiol. 10 (2), 259-271 (1993) 95020526 11 (bases 162129 to 164080) Grundy, F.J., Waters, D.A., Allen, S.H. and Henkin, T.M. Regulation of the <i>Bacillus subtilis</i> acetate kinase gene by CcpA J. Bacteriol. 175 (22), 7348-7355 (1993) 94042910 12 (bases 7009 to 15526) Kiel, J.A., Boels, J.M., Beldman, G. and Venema, G. Glycogen in <i>Bacillus subtilis</i> : molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation Mol. Microbiol. 11 (1), 203-218 (1994) 94195107 13 (bases 196487 to 200620) Jin, S. and Sonenshein, A.L. Identification of two distinct <i>Bacillus subtilis</i> citrate genes J. Bacteriol. 176 (15), 4669-4679 (1994) 94321340 14 (bases 16985 to 19588) Abe, A., Koide, H., Kohno, T. and Watabe, K. A <i>Bacillus subtilis</i> spore coat polypeptide gene, cots Microbiology 141 (Pt 6), 1433-1442 (1995) 95400496 15 (bases 131934 to 133970) Bolotin, A., Khazak, V., Stoyanova, N., Ratmanova, K., Yomanta, Y. and Kozlov, Y. Identical amino acid sequence of the <i>aroA(G)</i> gene products of <i>Bacillus subtilis</i> 168 and <i>B. subtilis</i> Marburg strain Microbiology 141 (Pt 9), 2219-2222 (1995) 96118703 16 (bases 25258 to 31212) Rowland, B., Hill, K., Miller, P., Driscoll, J. and Taber, H. Structural organization of a <i>Bacillus subtilis</i> operon encoding menaquinone biosynthetic enzymes Gene 167 (1-2), 105-109 (1993) 96144257
--	--

REFERENCE	AUTHORS	TITLE
17 (bases 196487 to 200620)	Jin, S., De Jesus-Berrios, M. and Sonenshein, A.L.	A <i>Bacillus subtilis</i> malate dehydrogenase gene
JOURNAL	J. Bacteriol.	178 (2), 560-563 (1996)
MEDLINE		96134995
REFERENCE		18 (bases 129888 to 132207)
AUTHORS	Varon, D., Brody, M.S. and Price, C.W.	<i>Bacillus subtilis</i> operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H
JOURNAL	Mol. Microbiol.	20 (2), 339-350 (1996)
MEDLINE		96310371
REFERENCE		19 (bases 81540 to 91690)
AUTHORS	Bower, S., Perkins, J.B., Yocum, R.R., Howitt, C.L., Rahaim, P. and Pero, J.	
TITLE		Cloning, sequencing, and characterization of the <i>Bacillus subtilis</i> biotin biosynthetic operon
JOURNAL	J. Bacteriol.	178 (14), 4122-4130 (1996)
MEDLINE		96312354
REFERENCE		20 (bases 49093 to 51682)
AUTHORS	Yocum, R.R., Perkins, J.B., Howitt, C.L. and Pero, J.	
TITLE		Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from <i>Bacillus subtilis</i>
JOURNAL	J. Bacteriol.	178 (15), 4604-4610 (1996)
MEDLINE		96345628
REFERENCE		21 (bases 100760 to 102298)
AUTHORS	Kappes, R.M., Kempf, B. and Bremer, E.	
TITLE		Three transport systems for the osmoprotectant glycine betaine operate in <i>Bacillus subtilis</i> : characterization of Opud J Bacteriol. 178 (17), 5071-5079 (1996)
JOURNAL		96359357
MEDLINE		22 (bases 1 to 220060)
REFERENCE		Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, S.D.
AUTHORS		Sequencing and functional annotation of the <i>Bacillus subtilis</i> genes in the 200 kb rnm-dnaB region
TITLE		Microbiology 143 (Pt 11), 3431-3441 (1997)
JOURNAL		98048467
MEDLINE		23 (bases 1 to 220060)
REFERENCE		Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, S.D.
AUTHORS		Direct Submission
TITLE		Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
JOURNAL		location/Qualifiers
FEATURES		1..220060
Source		/organism="Bacillus subtilis"
		/db_xref="taxon:1423"
		485..2034
rrna		/product="16S ribosomal RNA"
		2203..5128
rrna		/product="23S ribosomal RNA"
		5185..5300
rrna		/product="5S ribosomal RNA"
		5322..5397
rrna		/note="codon recognized: GUA"
		/product="tRNA-Val"
		5430..5504
rrna		/note="codon recognized: ACA"
		/product="tRNA-Thr"
		5543..5618
rrna		/note="codon recognized: AAA"
		/product="tRNA-Lys"
		5629..5715
rrna		/note="codon recognized: CUG"
		/product="tRNA-Leu"
		5721..5795
rrna		/note="codon recognized: GGC"
		/product="tRNA-Gly"
		5810..5895
rrna		/note="codon recognized: UUA"
		/product="tRNA-Leu"
		5905..5981
trna		/note="codon recognized: CGU"
		/product="tRNA-Arg"

trNA	5997. .6073	/note="codon recognized: CCA"	
trNA	6079. .6151	/product="tRNA-Pro"	
trNA	6172. .6248	/note="codon recognized: AUG"	
trNA	6251. .6327	/product="tRNA-Met"	
trNA	6334. .6425	/note="codon recognized: AUG"	
trNA	6443. .6519	/product="tRNA-Ser"	
trNA	6531. .6607	/note="codon recognized: AUG"	
trNA	6531. .6607	/product="tRNA-Met"	
trNA	6531. .6607	/note="codon recognized: GAC"	
trNA	6531. .6607	/product="tRNA-Asp"	
Query Match	9.1%; Score 108.6; DB 1; Length 220060;		
Best Local Similarity	49.3%; Pred.No.3.3e-22;		
Matches 316; Conservative	0; Mismatches 319; Indels 6; Gaps 1		
QY	521	TAGCAGACGAGCGCAATACGATCATGTGTTTAAAGGACATGATGATTTGTTATTT	580
DB	26151	TGGAGATGACGACGACAAAGCTTGTGTTTGCATTTGCAACAAAGGCAAAAGCTTTGTG	26210
QY	581	CAGCAACACCGGACGGTTAGTGGCTATGTCAGTGTCAGATCGCTACGGCGCGGTG	640
DB	26211	GGCGCTCTCCGGAAAGACTGATCAAAAGAGAGCGGCGGCTGTCATGTCTTCTGCTGG	26270
QY	641	CTGGGAGCAAGCGCGGCGGAGCGATGGCGCTGACGATATCCGCTTAGGCGAAAGCTTGT	700
DB	26271	CAGGCTCCATTAAACGAGGCGTGAAATGAAGAAGACCGCGGATGAGGCTTTGAATAT	26330
QY	701	TAGCCAGTCAGAAAAACCGCATTTGAACATCATATGTCGTGCGCAAGTATCAGCACGCT	760
DB	26331	TAAAGATGAGAAAAACCGCTTTGAGCATGATATGTGTAGTATGATTCATTAATGCTT	26390
QY	761	TCCAAAGCTGACGAGCTGCTTAAAGTGGCGGCAATGCCAAGTTTACTCAAAATATAGC	820
DB	26391	TTGTATCAAGCTGTTTCAGAGGTTGAAAAGCCGATGAGACCTGTTTGTACAAAACAAAA	26450
QY	821	AGTTCAGCATTTGTACACACCAATTTACAGGGGACATTCGGCACATTTAAAGTGCAGC	880
DB	26451	GGCTTCAGCACTGTTTACGCCGATTTGCGACGTCGGCAGTCTGCTCGATTTTGG	26510
QY	881	CGATGTTGACCGCTTGCATTCACACACGACACTGGTGGCTCCACGTCGAAGCGGCC	940
DB	26511	ACTTATTTGAGAAAGCTGCACCCCACTCCCGGCTTGGGGATGCACCTCAGAAAAAGCTG	26570
QY	941	TGTATTTACATTCGACCCCAATGAGAACACGCTGCGCTTGTGTCAGGTCCTATTGGCT	1000
DB	26571	TTGACGTGATCAGGAGATTTGAACGATGTCCCGGCTGTGATGCGGCCCTTAATGGCT	26630
QY	1001	ATTTTACCGCAATATTAATGGGGAATTTGTGTGGCATTCGCTTCCATGTATGCAATC	1060
DB	26631	GGATTGACAGCCAGGATTAACGGGGAAATTCGGCGCTCCCTTGTTCAGGGCTCATCGAAG	26690
QY	1061	AAACGAGGACGAGCAACTTTATTGCTGTGGCGGAGTGTGGCTGACTCCGATGCGC	1120
DB	26691	GAA-----GCACAGCAGGCTGTTTGGCGGGTGGCGGTTTGTGGAAGACTCGGAACAA	26744
QY	1121	AACAGAAATATGAGAAACTGGTGTGAATTTGAACCCATG	1161
DB	26745	TCTCGAATATGAGAAACACAGATTTAAATTAAGCCATG	26785

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
PAX82644	4662 bp	DNA	BCT	08-JAN-1997		
P.aeruginosa pchd, pchc, pchb and pcha genes.						
X82644						
X82644.1	GI:1628425					
pcha gene; pchb gene; pchc gene; pchd gene.						
Pseudomonas aeruginosa.						
Pseudomonas aeruginosa.						
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;						
Pseudomonas.						
1 (bases 1 to 4662)						
Serino, L., Reilmann, C., Baur, H., Beyerle, M., Visca, P. and Haas, D.						
Structural genes for salicylate biosynthesis from chorismate in						
Pseudomonas aeruginosa						
Mol. Gen. Genet. 249 (2), 217-228 (1995)						
96086939						
2 (bases 1 to 4662)						
Serino, L., Reilmann, C., Visca, P., Beyerle, M., Chisa, V.D. and						
Haas, D.						
Biosynthesis of pyochelin and dihydroaeruginosic acid requires the						
iron-regulated pchDCA operon in Pseudomonas aeruginosa						
J. Bacteriol. 179 (1), 248-257 (1997)						
97136628						
3 (bases 1 to 4662)						
Serino, L.						
Direct Submission						
Submitted (14-NOV-1994) L. Serino, University of Lausanne, Lab de						
Biologie Microbienne, Batiment de Biologie, 1015 Lausanne-Dorigny,						
SWITZERLAND						
Revised by (41						
4 (bases 1 to 4662)						
Serino, L.						
Direct Submission						
Submitted (27-SEP-1996) L. Serino, University of Lausanne, Lab de						
Biologie Microbienne, Batiment de Biologie, 1015 Lausanne-Dorigny,						
SWITZERLAND						
On Oct 23, 1996 this sequence version replaced gi:929778.						
Location/Qualifiers						
1..4662						
/organism="Pseudomonas aeruginosa"						
/strain="PAO1"						
/db_xref="taxon:287"						
211..229						
/note="Fur binding site ('fur box')"						
join(256..260,265..1905)						
/gene="pchd"						
256..260						
/gene="pchd"						
265..1905						
/gene="pchd"						
/note="similarity to adenylate-forming, activating						
enzymes"						
/codon_start=1						
/transl_table=1						
/protein_id="CAA57966.1"						
/db_xref="gi:1628426"						
/db_xref="SPRMBL:p72175"						
/translation="MTSSPYPSAVDDPMPAFVBRYLDAIGHMODSPFALATSPRA						
AHPRIALCDDDRLSYADLORRCRLAAGLRQGLAGHPVYVHLDRNGIAFVETCSR						
LFQGLVRYVALPRAHROHETSGFRFEAKRYTGAERTIDGDPMPAKELIASCARR						
ALHGEAEAPLQALAPLYOADALEDCAAREADICFQISGTTGTPKIPRRAREYVLA						
NVRSAEVCGEDEHTVYLTGTPMAHNTLTCCPGVITGLASGAVVSRADPEPCFLAL						
IARRYVTHALVPLPLAMLMADQESRADLSLLDLVGGSRIGSSAAQREVPVIGCC						
LOQVYGAEGILICTRLDDPPEPRLVHNGRGLSPDDVRYVDARGREVPGEVELTV						
RGPRTGGYRRLPEHNAKFSADGFYTGCBSVSDKKGIVLYBERDDQINRGEKTN						
AEVEENLLIAHPOVHDATVYAMPSPDLGERCAVITROPAAALKTLKQYHACGLAA						
EKVPDRLELPAFPQTGIGIKSKDLRLRELEARA"						
1890..1894						
/gene="pchc"						
1890..1894						
/gene="pchc"						
1902..2654						
/gene="pchc"						
1902..2654						
CDS						

				/gene="pchc" /note="similarity to thioesterases" /codon_start=1 /transl_table=1 /protein_18="CAAS7967.1" /db_xref="GI:1628427" /translation="MSAAWVPEFLTPMRLPLACFPFHAGSASFESKSWERLPDID LLEAFYQPEREDFENPATRLEDLDADGALALNDFAPAPLALFGHSIGALALETALF LFPAALRHLPFNVAHPAHPHROGGALHRCDEAALLPDRRGASASELLPFADRLFTL PILRADVOAITETTRYRAOPIALCALDVLTGSHDEFVSAAEQMSDSKRIPARLRPP GGHTLYLSGRDAVEHLRLRLHPDALREVA"
gene				2641..2646 /gene="pchb" /gene="pchb" 2641..2646 /gene="pchb" 2651..2959 /gene="pchb" 2651..2959 /gene="pchb" /function="conversion of chorismate to salicylate (together with pcha)" /codon_start=1 /transl_table=1 /protein_id="CAAS7968.1" /db_xref="GI:929779" /db_xref="SWISS-PROT:O51507" /translation="MMKTPEDCGTGLDIRAIRIDLDIYOALGRMDVKASRPKA SEAIIPAPERVAAMLPERRAKAEENGDAFFVEGPLQIIHWITIAEIQIKWRTROGA "
gene				2956..4386 /gene="pchb" 2956..4386 /gene="pchb" /function="conversion of chorismate to salicylate (together with pchb)" /codon_start=1 /transl_table=1 /protein_id="CAAS7969.1" /db_xref="GI:929780" /db_xref="SWISS-PROT:O51508" /translation="MSRIAPLSOCIHALRGTFERAIGQAQLDRPYLVASFEDPLD PIQAFGAMDRCGFDPCLWEQPELFAFMGCALIELOGHEORFARIEBNMQLCADAV EGPLAPRLCGEPFRDPGRPREHQAPADASLMLAGITVLREGERYVLCOLAKGE DALALAVHCSAILRLROPARRRRSPSGQAODSAOSROEARVSDSVSRGR FGKYVALRTQARPGLDIEPMOVYIELRLLOHADQQLPACRGNACFCGASPERLVIRA GKALTJHALAGTIARGSDAQEDARLGQALDSADRHEHQLVVAIRTALEPFESELEI PDAGKYLARIVOHNLNPIRIRALADAGGILLRLQALHPRPVAGVGRSLADVTROHE GMDSGWYAAPLGMIDGNGENDFIVALISALLTPGRGYLFAGCGLVGDSEAHHYRBTG LKLSAMEBALSATIGGLEVPLORGVA"
BASE COUNT	714 a	1719 c	1560 g	669 t
ORIGIN				
Query Match		8.6%	Score 103.4;	DB 2; Length 4662;
Best Local Similarity		47.5%	Pred. No. 8,7e-21;	
Matches 345; Conservative		0;	Mismatches 376;	Indels 6; Gaps 1;
Oy	441 CTTAGCCAAAGTCGTTTTTGCTGGGCAACAGACCCTGCAGTTATCGACACGATTAGACT	500		
Db	3609 CTTTCGGCAAAGTCGTCGGCCCGCACGCCGCGCTCTCGGCGCATCGAGCCGTG	3668		
Oy	501 GGCAACAATTATTCCTGGGTTTAGCTGACGACGAGCGCAATTCGATATCATGTGCTTTAAACG	560		
Db	3669 GCAGGTCATCGAACACCTGCCTTCGCAACATGACGACGCCAGCTGTCCTGCGCCG	3728		
Oy	561 ACATGATGAATTGTTTATTTTCAGCAACACCGGAAGGTTAGTGCTATGTCAGGTGGTCA	620		
Db	3729 CGGCAACGCTTCCTTCCTCGGCGCCTCCCGGAACGCTGCTCCGCAATTCGCGCGCGCA	3788		
Oy	621 GATCGCTACGGGCGGCTCGCTGGGACAAAGCCGGCGCGGACGGATGGCGTCAGCATAT	680		
Db	3789 GGCACCTCACCCATCGCCCTGCGCGGGACCATGCGCCGGGCGGCGATGCGCCGGAAGAATGC	3848		





gene  
CDS  
/note="ORF\_ID:sl12789"  
/codon\_start=1  
/transl\_table=11  
/product="resolvase"  
/protein\_id="BA18003.1"  
/db\_xref="GI:1653087"  
/translation="MGQFREMMLERQREGVAKAEKTEAKKPISDGVPEQVYVTLVRRH  
GSKSVARELITQASVYRIIINLCSA"  
complement(4340..5218)  
/note="ORF\_ID:sl11428"  
/codon\_start=1  
/transl\_table=11  
/product="p3 protein"  
/protein\_id="BA18004.1"  
/db\_xref="GI:1653088"  
/translation="MESNPLTPIPLPLFLIMFGMGLTIKDFNRIWLEPKKAVAIGI  
LQAVLMPVWFGSLSPLESLPELAAGMIACPGGSTNVITYLKGKVALSTITL  
TASSIVLITPIPLVYVNAANYEMGEOPALDLPKTVLOIAVTIIPVSGMFIYFR  
LPIKIMWERNKMLSTFPLGIITYAILVKEENLADFLDVGCGYVTLNITLNGLSG  
GIAIAGLGLAKERSITTEVGIQNGTLAIATVTLNLPIMAIIPAIYSVMFLTSGH  
FGLGKAKIFRGLKST"  
5614..6606  
/note="ORF\_ID:sl1503"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA18005.1"  
/db\_xref="GI:1653089"  
/translation="MTISTPTDPSLITIMPAGCGTMAWGKILMGYOPAMNGELRRVFR  
EYVAGITPFPGDGSYGTGRLOGSEKLLGQSPQAYBEIANDKTCIATKLAPYRWRLT  
PQAMGAGASQQRQLRPIDLVOLHWSYANAPWEOPLKGLVOLCGRGAKLGSLS  
NGPRLKXADPEQNEGVSYLSVQYSLSTDPVSKLTGLKECEGLKILVYSD  
TLGLTGKPGGPPGPIRGILFFRLLPKIQLPIDLTQALAREKETMAOVYALNMCID  
CGTLPPIPKAKKQOVQDNLGALGQLDRGEMDELNNQAALPQTMVQNIHQSA"  
complement(6657..7907)  
/gene="hob"  
/complement(6657..7907)  
/note="ORF\_ID:sl11427"  
/gene="hob"  
/codon\_start=1  
/transl\_table=11  
/product="protease hnob"  
/protein\_id="BA18006.1"  
/db\_xref="GI:1653090"  
/translation="MAIHLSKSLGVAVLLFLFGALIGAAGGYLLSSGNHSSPDSR  
VNTSLOSLPAVENSYRSALPILTPRSADODENLFARAVQKIGPAAVYIDSEFRPA  
VSQGGPMGQPFRRFEGEEMPNPDPEQGTSGSILSSGEVLTANAHVEGASTYK  
VTLKGSYLEGVKMDITPMTDYAVYKVAENLPVYEIQSRLDPGERATATGCPGLI  
DNTVYVGLISALGRSSSEGVDPDKRVRFIQTDAIINPGNSGPIPLNKAQEVIGVATLI  
RADAGLGALPIIQTQYNAWENLFTKGMENHYGLIHVYVLTPEMTKQLRTSGELPAGG  
VYADGVLGIQVSPGSPAQAQGIAPGDIILVGGWGYTATDQEREVEVSQIGBPLAI  
AVKROKPOMAWVVRGPPEDLGQ"  
complement(8021..8491)  
/note="ORF\_ID:sl11426"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA18007.1"  
/db\_xref="GI:1653091"  
/translation="MKSLKRGIALSYMATGCGYGLTYSQAYAPGDGDTITVOEALD  
TPVDVDTIDQDQLITWICKNGSKSVAAVEVELGQWETVPDPDTMACEONIPEYDVG  
DATFECENATGKLTITVWLEBEGGKAEKMTSMNNLNANNMVCITHGOSNPFME"  
complement(8617..10470)  
/gene="pros"  
/complement(8617..10470)  
/gene="pros"  
/note="ORF\_ID:sl11425"  
/codon\_start=1  
/transl\_table=11  
/product="proline-trna ligase"  
/protein\_id="BA18008.1"  
/db\_xref="GI:1653092"  
/translation="MWDGVSVYHLFRLIMRLSQMLFVTLRDDPADAEIPSHKLLVVRGQ

```

FIRLRSGSLVLPDLIRVLEKXKRIYVEENMRATQAGEOELLPOLQSLSLKMSGRMPLT
KTESBGTGFMALRDLREPLDELGLGTHHEVYITLAKEMIKSINIFRPGCLDFRVDVDSGASQ
KPRFGLMGKREIFRIMKDYSPHSDSEASLAKTATGAMODANRNIIFRGLDLOIOTKRSRDEI
YGGSSGQEPFVLADAGDELDTLYIYDEIYASANEKAVLVSIRLQDQKPSFEGSYKQSGATP
IGGSSGQEPFVLADAGDELDTLYIYDEIYASANEKAVLVSIRLQDQKPSFEGSYKQSGATP
IATKMCQCDLDCDASNNVKKVLAIOAYVNSGKNAVLVLSIRGQOEYNEKLNELTKLATA
PHYSESTNLSTLYDEDEKQOQKMAKSLPLGLTYSIPGISDYSITKNOVADKELRLVDTQD
ADVLENTFNTGADVADIVYVATGANNSTFPLTKLVLVDRLKADLRDVRHPTDELTAAGQ
LEVEGHIGLQKRSQAVLQNGEYDTEGDEEGRKELDYMGQSGVSLAQAAOVSQSDMGII
MPTLAPAEVYIYVPAQNGDEQMOVAEOITYPEQAOEYVDLIDDPENACQYKPRDSEI
IGIPRYVYTGKSLKQOGLVEYKRAKDSVONLAIAYVPLTLMIAIKKTS"
10639..11235
/note="ORF_ID:s1r1505"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAAL8009.1"
/db_xref="GI:1653093"
/translation="MDPLASSSASPVNQAHYVNMETLLYEIDKQLRFFPKNLNRNVL
NLTEVETALNRLPEPLVASSVCQVEEQQRNVAARQVRSLETSSVVRALIAAVERDPLRSSS
EPIPELEEVNRGMEQLGIDQLRLKRLNPLVYKVEYVEMDNCPLLRQALTKLTAQV
PPTPEAIKRENSMLPEPPIRITPAPDNQIDQPKTNY"
1

```

Query Match	7.5%;	Score 89.8;	DB 2;	Length 143051;
Best Local Similarity	48.9%;	Pred. No. 2.4e-16;		
Matches 303;	Conservative 0;	Mismatches 307;	Indels 9;	Gaps 2;

Qy	556	AAGGACATGATGAATTGTTATTATTCAGCAACACCCGAGCGTTAGTGGCTATGTCAAGT	615
Db	128593	AATGGCCAGGGGGGATTTGTTTGTGTGGGGCTAGGCCAAGAATCTTAAGTTTGCAATAT	1285625
Qy	616	GGTGAGATCGGTACGGCGCGGTGGTGGGACAGCCGGCGGGAGCGAATGGCGGTGAC	675
Db	128653	CAACAGCTGGTGCACCGATGGCCCTAGCAGTTCCGGCTCTAGGATGTGCATGTCCAGGC	128712
Qy	676	GATATCGCGTTAGCGGAAGCGTGTGTAGCCAGTCAGAAAAACGGCATTTGAATCAATAT	735
Db	128713	GATCGCCAAATTGGGGGCAAGATTTATTCACATATCCCAAGGAATTTGGCGGAACATCAAGC	128772
Qy	736	GTCGGGCAATATTCAGCAGACGCTTGCAAGAGCGTGAAGCGTGCATTAAGTGGCGGCG	795
Db	128773	GTGTGGATTACCTCTCTCCAGCGTCTCGGGCATTTGGGCGCTTAAGCCCGCCAGGCTTCTCC	128833
Qy	796	ATGCCAAGTTTACTCAAAAAATATAGCAAGTTTCAGCATTTGTATACACCAATTAACAGGGAC	855
Db	128833	TTG---AAGTTGCTTAATTTAGGCCAATATTACAGCACTTCAGCACCAAAATTCAGGCCCT	128889
Qy	856	ATTGCGGACATTTTAAGTGTGTACCGCGATTTGTTACCGCTTGCAATCCAAACACAGACTG	915
Db	128930	CTGGCGGCCCAATCCATCCCTCGCCCTGGTGTGAACAATCCATCCACCCCGCGTGTG	128949
Qy	916	GGTGGCGCCCAAGCAGTGAAGCGCGCTGTATTATACATTGCGACCCGTGAAGACACCTCGT	975
Db	128950	GCCGGTGTCCCCCGTAGAGCATGTCCGGAAGATTTAATCCGTGGCCATGAACCTTTGACCGC	129005
Qy	976	GCGTTGTTTGCAGGTCCTATTGSCCTATTTTACCGCAGATTAATAGTGGGGAATTTGCGTT	1035
Db	129010	AATCTCTATGCTGCTCCCTGGTGGTGGTAGAGTAGTAAGTAAATGCTGAGTTATTTGTC	129065
Qy	1036	GGCATCGGTTCCATGTATGTGAATCAACGAGCAGGAGACCACTTATTTAGTGGTGGCC	1095
Db	129070	GGTATTCGTTT-----TGCCCTTCTGTCCCGTATATGAGCGACACTATACGCCGGGGCG	129123
Qy	1096	GGGATTTGTGGCTGCACTCCGATGCCAACAAGATATTAAGAAACATGGGTTGAATTTGAA	1155
Db	129124	GGCAATTGTGGCGGTTCCGATGCCCTCAAGAAGTGGCAGAAATTTGAGTCTCAAACTTCAA	129183
Qy	1156	CCCATGGCGGCAATTGTTAA 1174	
Db	129184	ACTCTTGCGGCTTCTTTGA 129202	

## RESULT

LOCUS	ECOEUC	1659 bp	DNA	BCP	25-JAN-1994
DEFINITION	E. coli isochorismate synthase (entc) gene, complete cds.				
ACCESSION	M24142				
VERSION	M24142.1 GI:341117				
KEYWORDS	entc gene; isochorismate synthase.				
SOURCE	Escherichia coli.				
ORGANISM	Escherichia coli				
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
AUTHORS	1 (bases 1 to 1659)				
TITLE	Ozenberger, B.A., Brickman, T.J. and McIntosh, M.A.				
JOURNAL	Nucleotide sequence of Escherichia coli isochorismate synthase gene entc and evolutionary relationship of isochorismate synthetase and other chorisimate-utilizing enzymes				
MEDLINE	J. Bacteriol. 171 (2), 775-783 (1989)				
FEATURES	89123153				
source	Location/Qualifiers				
gene	1..1659				
CDS	/organism="Escherichia coli"				
	/strain="K12"				
	/db_xref="taxon:562"				
	380..1555				
	/gene="entc"				
	380..1555				
	/gene="entc"				
	/EC_number="5.4.99.6"				
	/codon_start=1				
	/transl_table=11				
	/product="isochorismate synthase"				
	/protein_id="AA16100.1"				
	/db_xref="GI:450376"				
	/translation="MDTSLAEVQQTMTATLAPNRFMFMSYRFTTSGCTAREDEPAVNGSDSPFOOKLALAFADAKQGIKNPVWGAIPDPDPSLITPESQSPSROKQASRRPTROSINLVVERQAIPEQTPEQWVARAAATAPVDKVLSTLITDPAADISGVLEIRLIAONPVSNFHVLAAGVGLGASPELLRKDERSSPLPGSARRODPDVLREAGNRRLASEKDRHEHEITVQAMKEVLRSSSLHVPSPOLITPTLMHLAPPEEKANSQENALTLACLHPETPLLSGFPHQAIQVYALEPPEPLFGIVGNCDSRNGMNVVTTIRKAKIRENMORLFAFAGATVPASSPLGEMRETKVLTMLNFGH"				
BASE COUNT	379 a 443 c 454 g 383 t				
ORIGIN					
Query Match	6.3%; Score 76; DB 2; Length 1659;				
Best Local Similarity	46.6%; Pred. No. 2.8e-12;				
Matches 284; Conservative	0; Mismatches 320; Indels 6; Gaps 1;				
563	ATGATGAATTCGTTATTTACAGACACCGGAAAGCGTTAGTGGTATGACAGGTGTCAGA	622			
939	ATGGTGGCGTCTCTGCTGGGGGCGACCCGGGAATGCTGCTACGTAAAGACGGCGGCTT	998			
623	TGCGTACCGCGCGGCTCCTCTGGGACMACCGCGCGGAGCAGGATGCGCTACGATATCG	682			
999	TTAGCTTCATTCGTTACCGGTTCCGGGCGTCGTACACCGCGATGAATGCTCATTCGCG	1058			
683	CGTTAGCGGAAGCGTTGTTAGCCAGTCAGAAAAACCGCATTTGAACATCATATATGTCGTG	742			
1059	AAGCAGGTAAATCGTCTGCTGGCGTCAAGAAAAAGATCCGCAATGAACATGAATCGTGAATC	1118			
743	CAGATATACGACACACGCTTGCAAGACGTAAGAGAGTGCCTTAAAGTGGCGGCATGCCAA	802			
1119	AGGCGATGAAGAAGGTACTGCGGCAAGCAGCATGATGATTTACACGCTTCTTCTCCAC	1178			
803	GTTTACTCAAAAATTAAGCAACGTTGACGATTTTGTACACACCAATTAACAGGGACATTCGGG	862			
1179	AGCTGATATCAGCAGCGGACGCTGTGGCATCTTCGCAACATCCCTTTGAAGTAAAGGAAATT	1238			
863	CACATTTATAGTGTGACCGCATTTGTTGACCGCTTGATTCACAAACACAGCAGCTGGTGGCG	922			
1239	CGCAGAAAGAACGATGACTCTGGCGCTTCTGCTGATCCGACCCCGCGCTGACCGGTT	1298			
923	TGCCAGGTAAGCGGCCCTGTATTAATCAATTTGGACCCATGAGAAAGACACTGCTGCTGT	982			
1299	TCCCGCATTCAGGCGCGCAGCCAGGTTATTTGGTGTGAATCTGGAACCGTTTCAGCCGGAAGCT	1358			

OY	963	TTGCAGGTCCTATTGGCATTTTACC	GCGAATTAATGTGGCAATTTTGCTGGCATTCC	1042
Db	1359	TTGGGGGCAATTCGTGGTGTGTGACAGCAAGGTAACGCCAAATGGGGTGGACATTC		1418
OY	1043	GTTCATCATGTATGTGAATCAAAACGACGACGACGACGACATTTATTTGCTGGTGGCGGATTTG		1102
Db	1419	GCTGGCGCAAGCTGGCGGAAAATCATCGTGGC-----TCTGTTTCCGAGGCGGGGATTTG		1472
OY	1103	TGGCTACTCCGATCCGCAACAAGATAATGAAGAACAATGGGTTGAATAATTTGAACCATGC		1162
Db	1473	TGCCGCGCTCGTCACACCGTTGGGTAGTGGCGCGGAACAGCCTCAACCTTTCTACCATCT		1532
OY	1163	GGCAATGCTT	1172	
Db	1533	TGAACGTTTT	1542	
RESULT	9			
AEO00165				
DEFINITION	AE000165	12003 bp	DNA	BCT 12-NOV-1998
LOCUS	Escherichia coli K-12 MG1655	section 55 of 400 of the complete genome.		
VERSION	AE000165.U00096			
KEYWORDS	AE000165.1 GI:1786808			
SOURCE	Escherichia coli.			
ORGANISM	Escherichia coli Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
REFERENCE	1 (bases 1 to 12003)			
AUTHORS	Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J.J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.			
JOURNAL	The complete genome sequence of Escherichia coli K-12 Science 277 (5331), 1453-1474 (1997)			
MEDLINE	97426617			
REFERENCE	2 (bases 1 to 12003)			
AUTHORS	Blattner,F.R.			
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	3 (bases 1 to 12003)			
AUTHORS	Blattner,F.R.			
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	4 (bases 1 to 12003)			
AUTHORS	Plunkett,G. III.			
JOURNAL	Direct Submission Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markborov@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible ( <a href="http://cgsc.biology.yale.edu">http://cgsc.biology.yale.edu</a> ). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli			











574 TTTATTTCAGCAACCGGAGCGTTAGTGGCTATGTCAGGTGGTCAGATCGCTACGGCG 633

```

/codon_start=]

```





Best Local Similarity 47.7%; Pred. No. 1.5e-07;  
Matches 287; Conservative 0; Mismatches 297; Indels 18; Gaps 3;

```
OY 574 TTTATTTCAGCAACACCGGAGCGTTAGTGGCTATGTCAGTGGTGCATATCCCTACGGC 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 793 TTTCTTGCTCTTCACCGGACGGTATGCGCGCGCTGACAAACCGCTGCTACTGAA 852
OY 634 GCGGTGCTGGGACAAACCGCGCGGACGATGGCGCTGACGATATGCGGTAGGCGAA 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 853 GCGGTGGGCGGAGTACGAAATATATCTGATATAGCAAGCGCCAGCAGTATAGAGAG 912
OY 694 GCGTTGTTAGCCAGTACGAAAAACCGCATTTGAATATATGCTGGCGCAAGTACAG 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 TGGCTGATGGGAGTATGATGAAAAACGCGCGAGACATGCTGGTGGAAATATCTGT 972
OY 754 ACAGCTTGCAAGACGTACGACGCTGAAGGTGCGCGGCATCCCAAGTTACTCAA 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 973 CAACGATTACAGCGCGGATACCGACGCTGATGTTTACCGCGCAGTACTGTGCGT 1032
OY 814 AATAAGCAAGTTACGATTTGTACACACCAATTTACAGGGGACATTTAGT 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1033 AAAGTGCAGCATCTTCCGCGCTGTATCTGACTTCACTCAACAAAGCGGATATGTATC 1092
OY 874 GTGACCGGATTTGACCGCTTGCCATCCAAACGACGACGCTGGGCTCCACAGTGA 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1093 TG-----TTTACATCAGTTGACGACGACGCGAGCTTGCTGCTTACCGCGCAT 1143
OY 934 GCGGCGCTGTATTCATTTGCGACCCATGAGAAACACTCGTGGCTTTGTCAGTCT 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 CTGGCGGACAGTTTATGCGCGCTGACGACCGTTCACCGGAATGATGATGATGAT 1203
OY 994 ATTGCTATTTTACCGCAGATATATGAGGGAATTTGTGTGGCATCGTTCATGAT 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 GCGGCTATCTCTCTATTTACAAACAAAGCG-----AATTTCGCTTTCCCTCGCTCAGC 1256
OY 1054 GTGAATCAACGCGACGACGACAACTTATTTGCTGCTGCGGATGTGGCTGACTCC 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 AAAAATTTGCGCATGTCGTCGCA--TTATATGCTGCGCGGCGGATTTGCTGTTCC 1314
OY 1114 GATGGCAACAAAGATATGAAAGAACTGGTTGAATTTGAACCATGCGCAATTTGA 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1315 GACCCGAGCAAGAGTGCAGAGAAATCGAACAAACGCGCGGCTGCTGACTTATTA 1374
OY 1174 AA 1175
Db 1375 CA 1376
```

RESULT 14  
AF078080 1973 bp mRNA PLN 23-DEC-1998  
LOCUS AF078080  
DEFINITION Arabidopsis thaliana isochorismate synthase mRNA, complete cds.  
ACCESSION AF078080  
VERSION AF078080.1 GI:3348076  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS Meng, H., Pullman, G. S. and Peter, G. F.  
TITLE Cloning of a Plant Isochorismate Synthase (Accession No. AF078080)  
JOURNAL (FEBS 214)  
REFERENCE Plant Physiol. 118 (4), 1536 (1998)  
AUTHORS 2 (bases 1 to 1973)  
TITLE Direct Submission  
JOURNAL Submitted (14-JUL-1998) Forest Biology, Institute of Paper Science  
FEATURES and Technology, 500 10th St, Atlanta, GA 30318, USA  
SOURCE 1. 1973  
/organism="Arabidopsis thaliana"

CDS

```
/db_xref="taxon:3702"  
217..1728  
/function="catalyzes chorismic acid to isochorismic acid"  
/note="1csi"  
/codon_start=1  
/product="Isochorismate synthase"  
/protein_id="AAC97926.1"  
/db_xref="GI:3348076"  
translation="MTAVLSPAAMTERLISAVSLKSGPSFGVVRIOVPDIOIG  
AIDMLQANETOPRCFPRSBRSDVSRPDLILDLANENGNGNNGVSSDRINVSATGIG  
SAVEFRDIDPESHDDMSIRRELSSTPLIRAYGMRDPNGKIAVEPRGATYFISV  
POVEFNEFGSSMLAAI IAWDELISLENAIEALQETMLQVSSVVKLRNLSGVSV  
LSKNHVPKGAFFPAVEKALEIMKSNSEPLKVVILANRSRIITDIDIDIAMLQOR  
EGHDAYOFCLOPPCAPFAGTPEPRLEORTLGVSEALATPRRAASADMEIERD  
LTSRKDLESTREYRENTREKNTICDRVYVVKPKQTKLARNOHIXSOLAGRLTERD  
DEKILALHETPAVCGIPAEERALLKEISEPRGMAGTGFGESESEFANGTIS  
ALVERGALIVAGTGVAGSDPSENNEDLKIISQFKSLIEYATTSLOAIN"  
BASE COUNT 530 a 392 c 460 g 584 t 7 others  
ORIGIN
```

Query Match 5.0%; Score 59.8; DB 7; Length 1973;  
Best Local Similarity 45.8%; Pred. No. 3.3e-07;  
Matches 205; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

```
OY 574 TTTATTTCAGCAACACCGGAGCGTTAGTGGCTATGTCAGTGGTGCATATCCCTACGGC 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1096 TTTATGGAACACCGCTGAAAGACTATTCAAAGACTCGTATAGTGTCTGCAGTAA 1155
OY 634 GCGGTGCTGGGACAAACCGCGCGGACGATGGCGCTGACGATATGCGGTAGGCGAA 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1156 GCTTGGCTGCAAACTAGGCTGTAGACTGCTTCTAGTCTCGATATGAGATAGAGCT 1215
OY 694 GCGTTGTTAGCCAGTACGAAAAACCGCATTTGAATATATGCTGGCGCAAGTACAG 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1216 GACTTACTAACCGTCCGAAAGACGACCTCGAGTTCTCTATGTCGAGAGAAATTAAGA 1275
OY 754 ACAGCTTGCAAGACGTACGACGCTGCTAAAGTGCGCGGCATCCCAAGTTACTCAA 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1276 GAAAGTTAAACNGTATATGTCACAGAGTTGTTGCAACCCCAAAAACCTGTAGAGAA 1335
OY 814 AATAAGCAAGTTACGATTTGTACACACCAATTTACAGGGGACATTTGCGCGCAATTTAAGT 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1336 CTTCGAAGAGTGCACACTATATATTCATATGCGAGGAGACATTCAGAAAGAAATGAT 1395
OY 874 GTGACCGGATTTGACCGCTTGATCCAAACACGACGACGATGGGCTGCCAGCTGAA 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1396 GAGTATAAATATTTGGCTGCTTCGATCCAACTCGAGCTGTTGTGGCTTCACAGCAA 1455
OY 934 GCGGCGCTGTATTCATTTGCGACCCATGAGAAACACTCGTGGCTTTGTCAGAGTCT 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1456 GAAGCAAGGCTTTGATTAAGAGATATGATATTCGATAGAGAAATGATATGCGGAGCT 1515
OY 994 ATTGGCTATTTTACCGCAATATAGTG 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1516 ATTGGATTTTGTGGCGAGAGAGAGTG 1543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15  
CRO6065 2078 bp mRNA PLN 20-OCT-1999  
LOCUS CRO6065  
DEFINITION Catharanthus roseus mRNA for isochorismate synthase.  
ACCESSION AJ006065  
VERSION AJ006065.1 GI:4127687  
KEYWORDS Isochorismate synthase.  
SOURCE Madagascar periwinkle.  
ORGANISM Catharanthus roseus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Asteridae; eusterids I; Gentianales; Apocynaceae;  
Catharanthus.  
REFERENCE 1 (bases 1 to 2078)  
AUTHORS van Tegelen, L.J., Moreno, P.R., Croes, A.F., Verpoorte, R. and

TITLE Willems, G.J.  
JOURNAL Purification and cDNA cloning of isochorismate synthase from  
MEDLINE elicited cell cultures of *Catharanthus roseus*  
REFERENCE Plant Physiol. 119 (2), 705-712 (1999)  
99137924  
AUTHORS 2 (bases 1 to 2078)  
van Tegelen, L.  
JOURNAL Direct Submission  
Submitted (08-MAY-1998) van Tegelen L., Department of Experimental  
Botany, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen,  
THE NETHERLANDS

## FEATURES

## source

Location/Qualifiers

1..2078

/organism="Catharanthus roseus"

/strain="G. Don (Madagascar periwinkle)"

/db\_xref="taxon:4058"

31..1773

/EC\_number="5.4.99.6"

/codon\_start=1

/product="isochorismate synthase"

/protein\_id="CA06837.1"

/db\_xref="GI:412768"

/translation="MASTGHCVAFHTDLSRKSSFFSNNSLSFRKSTNIYTRK  
KYRCSTLSMNGCMGDBRPVGTETETRTIPAVSPALAMERLSAVANLSTLPSAQ  
SGIIRLEVPFIEHIALDMLISQDKNLIPRCYBSGRSQVTFSDTSLNLRNGSAA  
NGHLQRISTSSDDKNLVSAVGSVLRSPNPFESDMLSKRFLSKNCPILRAYGA  
IRPDARPHIAPEWKAFFSFEVYPOVEDELHSSMIAATVAMDNALSLTYQOALVAL  
QTMEOVSTTSKLRQDVSHTSIVSKANIPRTSMDLTINRYLEEIGNKYSPLTKVL  
ARRSOVITSDIDPLAMLSFRKADKDAYOCLQHPAFAFIGNTEQLEGRDOLTFE  
SEALATATRARGESDLSLOMAHDLFSSPDNHEFAIVRENTROKLDATCTSVETPMK  
SVRKIKRIQHLIYARFRAGRLRSEDDFKLLSLHPTPAYCVGPFMEDARKFIANEMFDR  
GLIVGVGFEGASQDSFVSIGIRSLALIGKDALIYAGLVVEGSDPALEWDELKAS  
QFMKIMKLEAPALK"

BASE COUNT 595 a 423 c 418 g 642 t  
ORIGIN

## Query Match

Best Local Similarity 4.8%; Score 57; DB 8; Length 2078;

Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

QY 806 TACTCAAAATAATGACAGTTGACAGTTGTACACACCAATTAAGGGACATTGGCGCAC 865  
DB 1385 TAAGAAAGCTTAAGAAATCAACATCTTAAGCTCGATTGCGAGCAGATTACGCTCTG 1444  
QY 866 ATTTAAGTGTACCGGATTTGACCGCTTGATCCACACGACGACTGGTGCGCTCC 925  
DB 1445 AAGATGATGATTCAGATTTTGTCTCCCTTCATCTCAGCTGTTGTGGGTTTC 1504  
QY 926 CACGTGAAGCGCCCTGTATTACATTGCGACCCATGAGAAGACACCTGCGCTTGTG 985  
DB 1505 CTATGGAAGATGACGGAATTAATTATTCGGAATAATGATTGTTGACGAGATTATACG 1564  
QY 986 CAGGTCCATTGGCTATTTTACCGCAGATAATAGTGGGAATTGTGGTGGCATCCGT 1045  
DB 1565 CTGGCCCTGTTGTTCTTTGAGAGAGCTCAGAGTATTTTCTGTTGGAATTAAGATCTG 1624  
QY 1046 CCATGTATGTGAATCAAGCAGCAGCACTTATTGCTGTCGCGGATGTGG 1105  
DB 1625 CC-----TTGATTTGGAAGAGATCCCGGTGCTAATATATGCGGCGCTTGGGCTGTAG 1678  
QY 1106 CTGACTCCGATGCGCAACAAGATATGAAGAAACTGGGTTGAA 1148  
DB 1679 AAGGAAGTATCCAGCTCTAGAAATGCGAGAACTAGAGCTCAA 1721

Search completed: November 25, 2000, 01:55:45  
Job time: 2265 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2000, 01:03:28 ; Search time 94.96 Seconds  
(without alignments)  
4735.344 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197  
Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTCGATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

N.Geneseq.36:\*

1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1197	100.0	1197	19	VA1196	Human neutral sph
c	130.6	10.9	4328	18	V41590	Staphylococcus aur
3	102.2	8.5	11764	20	X12985	Enterococcus faeca
4	76	6.3	1659	20	Z23756	E. coli entc DNA.
5	57	4.8	2078	20	Z23758	C. roseus ICS DNA.
c	42.2	3.5	10357	19	V52324	Streptococcus pneu
7	41.6	3.5	5057	20	Z23757	Streptococcus pneu
8	36.2	3.0	1664976	19	V21209	P. fluorescens DNA
c	33.8	2.8	1167	19	Z96384	Mechanococcus jann
10	32.4	2.7	3647	18	V74332	S. pneumoniae deri
11	31.6	2.6	16870	20	X13035	Staphylococcus aur
12	30.8	2.6	2032	13	Q29937	Enterococcus faeca
						PEP4 gene encoding

	13	30.8	2.6	3640	20	X88786	Cyclorella cryptic
	14	30.6	2.6	2161	20	Z30444	Tobacco anthranila
C	15	30.4	2.5	13926	19	V52138	Streptococcus pneu
	16	30.2	2.5	231	18	T68178	H. pylori cell env
	17	30.2	2.5	648	19	X30483	H. pylori cellular
C	18	30	2.5	7986	20	Z23937	T. versicolor lacc
	19	29.8	2.5	1275	20	E12105	Neisseria meningit
	20	29.8	2.5	1575	20	E12105	Neisseria meningit
C	21	29.6	2.5	1478	20	X07180	Corn threonine dea
	22	29.6	2.5	21126	20	X12960	Enterococcus faeca
	23	29.6	2.5	21170	20	X20535	Polynucleotide seq
C	24	29.4	2.5	1383	9	N80872	Nucleotide sequenc
	25	29.4	2.5	1419	16	O89692	pepr. Lactococcus
	26	29.4	2.5	1650	20	Z30446	Tobacco anthranila
	27	29.4	2.5	2380	19	V68834	Zea mays soluble s
	28	29.4	2.5	11597	20	X13159	Enterococcus faeca
C	29	29.4	2.5	13085	20	X13316	Enterococcus faeca
	30	29.4	2.5	24417	18	T97221	Pseudomonas aerugi
	31	29.2	2.4	540	21	A05526	Streptococcus pneu
C	32	29.2	2.4	554	21	Z27314	Human ion transpor
C	33	29.2	2.4	701	20	X60243	Domain IV sequence
C	34	29.2	2.4	2573	18	T77804	CDNA encoding vari
C	35	29.2	2.4	3023	18	T67241	Pasteurella haemol
C	36	29.2	2.4	3336	21	Z46368	Haemorrhagic enter
C	37	29.2	2.4	4503	20	X13086	Enterococcus faeca
	38	29.2	2.4	6236	19	X75759	Human neurofilamen
C	39	29.2	2.4	6524	18	T77803	CDNA encoding wild
C	40	29.2	2.4	6527	18	T77806	CDNA encoding vari
C	41	29.2	2.4	7052	18	T77805	CDNA encoding vari
C	42	29.2	2.4	10254	19	V52162	Streptococcus pneu
C	43	29.2	2.4	26270	21	Z46355	Haemorrhagic enter
44	29	2.4	247	19	V27388	Streptococcus pneu	
45	29	2.4	1119	12	Q12159	Truncated Met-stre	

## ALIGNMENTS

RESULT 1	
VA1196	VA1196 standard; CDNA; 1197 BP.
XX	XX
AC	VA1196;
XX	XX
DT	20-OCT-1998 (first entry)
XX	XX
DE	Human neutral sphingomyelinase encoding CDNA.
XX	XX
KW	Human: neutral sphingomyelinase; N-SMase; diagnosis; Crohn's disease;
KW	obesity; diabetes; Alzheimer's disease; ss.
XX	XX
OS	Homo sapiens.
XX	XX
FH	Key
FT	Location/Qualifiers
FT	1.1197
FT	/tag= a
FT	/product= "neutral sphingomyelinase"
FT	/transl_except= (pos:175..180,aa:Val)
FT	/note= "appears to encode 2 Val residues but only one"
XX	is given in the amino acid sequence (W69162)."
PN	W09828445-Al.
XX	XX
PD	02-JUL-1998.
XX	XX
PF	23-DEC-1997;
XX	97MO-US24051.
XX	XX
PR	24-DEC-1996;
XX	96US-0774104.
XX	XX
PA	(UYJO ) UNIV JOHNS HOPKINS.
XX	XX
PI	Chatterjee S;
XX	XX

DR WPI; 1998-377673/32.  
DR P-PSDB; W69162.  
XX Human neutral sphingomyelinase - used to, e.g. treat N-Smase  
PT associated disorders, e.g. Crohn's disease, obesity, diabetes, and  
PT Alzheimer's disease  
XX  
PS Claim 2; Page 29; 47pp; English.  
XX  
CC The present sequence encodes human neutral sphingomyelinase (N-Smase).  
CC A host cell containing a vector comprising a nucleotide sequence  
CC encoding N-Smase can be used to produce N-Smase. N-Smase can be used in  
CC a method for identifying a compound useful in the diagnosis or treatment  
CC of a human neutral sphingomyelinase related disorder. N-Smase, and a  
CC nucleotide sequence encoding N-Smase, can be used for modulating N-Smase  
CC activity, and for treating a disorder associated with N-Smase. The  
CC N-Smase disorders that can be treated with the nucleotide sequence or  
CC N-Smase, is an inflammatory disorder, arthritis, osteoarthritis, Crohn's  
CC disease, obesity, diabetes, cirrhosis, susceptible tumours, central  
CC nervous system disorder, vascular stenosis, arterial occlusion arising  
CC from plaque formation, cardiac disease where IV dysfunction occurs,  
CC hypercholesterolemia, cholesterol ester storage disorder, renal failure,  
CC HIV infection, depression, schizophrenia, neurodegeneration, and  
CC Alzheimer's disease. An antibody against N-Smase can be used to reduce  
CC tumour necrosis factor alpha (TNF-alpha) induced apoptosis of mammalian  
CC cells.  
XX  
SQ Sequence 1197 BP; 314 A; 267 C; 319 G; 297 T; 0 other;

Query Match 100.0%; Score 1197; DB 19; Length 1197;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGACATATCAAGAACCGCGCGTGGCTCAAAAGCACTTACACACTCTATCG 60  
DB 1 atgtatgacatatacgaacgcgcgcttggctcaaaagcacttacagcaactctatcg 60  
QY 61 GCATTGGAACACATGAAATTTGGCGCTTACCTTGGAGACCCGGGATGATTAAGT 120  
DB 61 gcaattggaacacatgaaatttggcgcttaccttggagacccgggatgattaaagt 120  
QY 121 TTTGGCATTTGGCGCAATTCGTACGCAAAAAGCGCTCAGCATTAACAAGTGGGTTGT 180  
DB 121 ttggcatattggcgcaatctgacgtacgcaaaaagcgctcagcattacaagtgcggtt 180  
QY 181 TTTGGTGGCGAGTATTTGATGACAGAGATCCGCGAGTGTGAATTTATGGCGGTTT 240  
DB 181 tttgggtggcgagtattgattgacagagatccgcgagtggtgaatttatggcggttt 240  
QY 241 TGGTTTGCCCGAAGTATGATGACATCCGCGAGATAAATACAGTTGGAGTCAGAT 300  
DB 241 tggtttgcccgaaagtatgattgacatccgcgagataaatacagttcgatcagat 300  
QY 301 ACCGTATCTGATTTTACAGCTGGCTGGCGAGTTGCTGCAAAAACAGCAAAATACGGTG 360  
DB 301 accgtatctgattttacagctggctggcgagttgctgcaaaaacagcaaaatacgg 360  
QY 361 ACCACTAGTATGTGACAGATGAAGTGTGATCGAACGACAGACAAATTTGATTAT 420  
DB 361 accactagtatgtgacagatgaagtgtgattgatacgaacgacagaaatattgat 420  
QY 421 ACCTTAGCATCGATCAAACTTAGCCCAAGTCGTTTGGCGGACACAGCCCTGCAG 480  
DB 421 accttagcatcgatcaaaacttagcccaagtcgtttggcggaacacagccctgcag 480  
QY 481 TTATCGACAGCTTACGACTGGACAAATTTTCGTGCGTTAGCTGACAGCGAATACG 540  
DB 481 ttatcgacagctttagcactggacaaattttcgtgctggttagctgacagcgaa 540  
QY 541 TATCATGTGCTTTTAAAGCAGATGATGAATGTTTATTTTCAGCAACACCGGAACGGTTA 600  
DB 541 tatcatgtgcttttaaagcagatgatgaattgtttatttcagcaaacacggaa 600

QY 601 GTGCTATGTCAGTGTGTCAGATCGCTACGGCGCGGTGCTGGGACAAGCCGCGCGG 660  
DB 601 gtgctatgtcagtggtgtcagatcgctacggcgcggtgctgggacaagccgcgcg 660  
QY 661 ACGGATGGCGCTGACGATATCGCTTAGCGGAAGCGTTGTTAGCCAGTCAGAAAAACCG 720  
DB 661 acggatggcgctgacgatatcgcttagcggaagcggttgttagccagtcagaaaa 720  
QY 721 ATTGAACATCATATGCTGTGGCAGTATCAGACACCGTTGCAGAGCTGACGACGTGC 780  
DB 721 attgaacatcatatgctgtggcagtatcagacaccgttgcagagctgacgacgtgc 780  
QY 781 CTTAAGGTCGCGCGCATGCGCAAGTTTACTCAAAAATTAAGCAAGTTCAGCATTTTGAACA 840  
DB 781 cttaaagtgcgcgcatgcgcaagtttactcaaaaatgaagcaagttcagcattttga 840  
QY 841 CCAATTACAGGAGACATTTGCGGACATTTTAAGTGTAGCCGCGATTTGACCGCTTGAT 900  
DB 841 ccaattacagagacatttgcggaacattttaagtgtagccgcatttgaccgcttgat 900  
QY 901 CCAACACGACACTGGGTGGGTGCCAGTGAAGCGGCGCTGTATTACATTCGACCCAT 960  
DB 901 ccaacacgacactgggtgggtgccagtgaagcgcgctgtattacattcgacccat 960  
QY 961 GAGAAGACACCTCGTGGCTGTTGTCAGTCCATTATGGCTATTATACGACATTAATAGT 1020  
DB 961 gagaagacacctcgtggctgttgcagtccattatggctattataccgacataatagt 1020  
QY 1021 GGGGAATTTGTGTGGATCCGTTCCATGTATGTAAATCAAAAGCAGAGCACT 1080  
DB 1021 ggggaatttgtgtggatccggttcctcatgtatgtaaatcaaaagcagacgagcaact 1080  
QY 1081 TTAATTCGTGTCGCGGAGATTCGTGCTGATCCGATGGCGCAACAGTAATTAACAAGT 1140  
DB 1081 ttaattcgtgtcgcggagattcgtgctgatccgatggcgcaacagtaattacaag 1140  
QY 1141 GGGTTGAATTTGAACCCATGGCGCAATTTGTTAAAGCACTACAAATCATGTCGAATGA 1197  
DB 1141 gggttgaatttgaacccatggcgcaatttgttaagcaactacaatcatgctgaatga 1197

RESULT 2  
V74590/c  
ID V74590 standard; DNA: 4328 BP.  
XX  
AC V74590;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE Staphylococcus aureus contig SEQ ID #279.  
XX  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
XX  
OS Staphylococcus aureus.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1081..1140  
FT /\*tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature 2881..2940  
FT /\*tag= b  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

XX EP786519-A2.  
 PN  
 PD 30-JUL-1997.  
 XX  
 PF 07-JAN-1997; 97EP-0100117.  
 XX  
 PR 05-JAN-1996; 96US-0009861.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;  
 DR WPI: 1997-374922/35.  
 XX  
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 PT stored on computer readable medium and used in the production of  
 PT anti-*S.aureus* vaccines  
 XX  
 PS Claim 1; Page 1101-1103; 3271pp; English.  
 XX  
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the *S.aureus* DNA sequences allows regative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against *S.aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the *S.aureus* DNA sequences contained on the  
 CC computer readable medium.  
 XX  
 XX Sequence 4328 BP: 1282 A; 776 C; 626 G; 1523 T; 121 other;

	Query Match	10.9%;	Score 130.6;	DB 18;	Length 4328;	
	Best Local Similarity	47.4%;	Pred. No. 4.7e-33;			
	Matches 354; Conservative	0;	Mismatches 384;	Indels	9;	Gaps
OY	451 GTCCGTTTTGTGCAGCACACGACCCTGCAGTATCCGACACGTTACGACTGGCACAATT	510				
Db	2942 GTNNN	2883				
OY	511 ATTCGCGCTTAGCTGAGCAGCGGAATACGTATCATGTGTTTTAAAGCAGCATGATGA	570				
Db	2882 NNGCAAGCATATTTAAAGAAAAAACAAGTATATATCTTCTTGAAATCACAGAATTCT	2823				
OY	571 TTGTTATTTCAGCACACCAGCGTTAGTGGCTATGTCAAGTGTCAGATCGCTACG	630				
Db	2832 ATATCTTTTTCACAAACACCTGMACTGAATAATTAAGATCAATAATAATATATCGACT	2763				
OY	631 GCGGGGTCGCTGGGACAAAGCCGCCGGGAGCGATGCGCTGACAGATATCCGTTAGGC	690				
Db	2762 AAAGCTGAGAGGCTACAAATTAACGTTTCACAAAGATGAGGAGGAATACAAAAATGTT	2703				
OY	691 GAAGCGTGTTAGCCAGCAGCAAAAACCGCATTTGAACATCATTTGCTGGCAGATATC	750				
Db	2702 GAAGCATTTTAAAAAGATTAATTAACACTTAATCGAACATCGATTTGTTGTTGACGATTT	2643				
OY	751 ACGACACCGCTTGACAGAGCGTAGCAGCGTGTAAAGTGCCGCCCATCCCAAGTTTACTC	810				
Db	2642 TTACATGATATTTAAACCTTATATACCTGAATTTACATTTGATTAAGACCGCTTAATTTCTA	2583				
OY	811 AAAAATAGCAAGTTGACAGATTTGTACACACCACAATTACAGGGGACATTTGGCGCACATTTA	870				

Db	2582	AAAAATGATCATTTATATCAGCTGTACACTGAAATTAAGAGGGCCACTGAAAGATGATTCG	2523
QY	871	ACTGTGACCGCGATTTGTGACCGCTTGATCCACACAGCACTGGTGCGCTCCACGT	930
Db	2522	TATATTAGTTTAAATGATAA---TTTACATCCAAACACTGCTTGTAGTGCGCTATCCAAA	2466
QY	931	GAAGGGGCCGCTATATACATTTGCCAGCCATGAGACACACTCGTGCTTTGGCAGCT	990
Db	2465	GAATTTGCATGATTTTATTTATGAGAGAAAGAAATTTGGTGTACGCGAGAAATATATGTGCA	2406
QY	991	CCATATTTGCTATTTACCGCAGATTAATATAGTGAGCAATTTGGTGGCATCCGTTCCATG	1050
Db	2405	CCGGTTGGCTATATATGATATATATACATGATGTGTGAATTTATTTGTTCGCAATTCGTTG	2346
QY	1051	TATGTGAATCAACAGCAGCAGCAGCAACTTATTTTGTGTGTCGCGGATTTGTGGCTGAC	1110
Db	2345	CTTATTAGAAAGC-----ACAACCACTTATTTTGTCTGGGTGAGCATTTGTAAGAT	2292
QY	1111	TCCGATGGCGAACAAGATATGAAAGAACTGGGTTGAAATTTGAAACCATCGGCAATTG	1170
Db	2291	TCTGATCCAGAAATGGAATTTGGCAGAAAGCAACTTAAGTTACACACTATGATGATGCA	2232
QY	1171	TTAAAGACTACATCATGTGCATGA	1197
Db	2231	TTAGGAGTGCATATGATGGAATCA	2205
RESULT	3		
XX	X12985		
XX	X12985	standard; DNA; 11764 BP.	
XX	X12985;		
DT	19-MAR-1999	(first entry)	
XX			
DE	Enterococcus faecalis genome contig SEQ ID NO:48.		
XX			
KM	Enterococcus faecalis; config; detection: Enterococcal infection;		
KW	vaccine; attenuation; computer readable medium; ds.		
OS	Enterococcus faecalis.		
XX			
PN	W09850555-A2.		
XX			
PD	12-NOV-1998.		
XX			
PE	04-MAY-1998; 98WO-US08985.		
XX			
PR	14-NOV-1997; 97US-0066009.		
PR	06-MAY-1997; 97US-0044031.		
XX	16-MAY-1997; 97US-0046655.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Barash SC, Dillon PJ, Kunsch CA;		
XX			
DR	WPI; 1999-045171/04.		
XX			
PT	New isolated Enterococcus faecalis polynucleotides and polypeptides		
PT	used to develop products for the detection of Enterococcus and for		
PT	use in vaccines for prevention or attenuation of Enterococcus		
PT	infection.		
XX			
PS	Claim 1; Page 438-444; 2084pp; English.		
XX			
CC	A computer readable medium has been developed which has recorded on it		
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome		
CC	X12938 to X13919 represent these nucleotide sequences which are primary		
CC	nucleotide sequences, also known as configs. The computer-based system		
CC	can identify fragments of the Enterococcus faecalis genome with		
CC	commercial importance. The products can be used to detect the presence		
CC	of Enterococcus faecalis in samples. They can also be used for		
CC	diagnosing Enterococcal infection in an animal and monitoring		

XX Enterococcus faecalis genome contig SEQ ID NO:48.  
DE  
XX Enterococcus faecalis; contig; detection; Enterococcal infection;  
XX vaccine; attenuation; computer readable medium; ds.  
KW  
XX Enterococcus faecalis.  
OS  
XX WO9850555-A2.  
PN  
XX  
PD 12-NOV-1998.  
XX  
XX 04-MAY-1998; 98WO-US08985.  
PF  
XX 14-NOV-1997; 97US-0066009.  
PR 06-MAY-1997; 97US-0044031.  
PR 16-MAY-1997; 97US-0046655.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Barash SC, Dillon PJ, Kunsch CA;  
PI  
XX WPI; 1999-045171/04.  
DR  
XX  
XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PT  
XX  
XX Claim 1; Page 438-444; 2084pp; English.  
PS  
XX  
XX A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome  
CC X1938 to X1919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring

progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, in vivo or in vitro. In particular the CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.

Sequence 11764 BP; 3614 A; 2155 C; 2554 G; 3428 T; 13 other;

Query Match 8.5%; Score 102.2; DB 20; Length 11764;  
Best Local Similarity 49.0%; Pred. No. 2,6e-23;  
Matches 365; Conservative 0; Mismatches 368; Indels 12; Gaps 3;

```
423 CTATGACATCATCAACCTTACGCAAACTGCTTTGGTGGGACACAGACCTGCGAGTT 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4727 cctaagagaagatgccccttaaaaaagttgttgcgtcgacaactcgtcgtgagcg 4786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 ATCCGACACGTTACGACTGGACAAATTTATGTCGGTTA--GCTGAGCAGCGAATAC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4787 tccgtaaaaaatcgaagtaaccagtgctcgttaatttaigtgcaacaacaatatc 4846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
540 GATATCATGTGTTTAAACGACATGATGATTTATTATTCAGCAACACGGAGGTT 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4847 gtaattgttcttagaagccaagaacaagttcattcgtgtgcaaccagaaagact 4906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 ACTGGCTATGAGTGTGACATGCTACGGCGGCGGCTGCTGGGACAAAGCGGCGG 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4907 tctgttgtaactaaagaacgtttgcgacgcgtgtatgcgcagaaacataacgg 4966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
660 GACGATGGCGCTGACGATATGCGCTTAGGCGCAAGCGTTGTTAGCCAGTCAGAAAA 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4967 tcaaacacagaagaagaacaaacttagagagacaactttacaagacgcaaaaatac 5026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
720 CATTGAACATCATATGCTGTGGCAAGTATACAGACACGCTTCAGACAGTGCAGAC 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5027 cggagaacacacaatgltgtgaaagtttagcgaagaagttgycgaagatgcagact 5086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780 GGTAAAGTGGCGGCGCATGCTTACTGCANAAATTAAGCAGTTGACATTTGTACAC 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5087 tgaaaacagatltcaagcaaccatatttggaaaatcgaagatgtaacaacattatgt 5146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
840 ACCAATTACAGGGGACATTGCGGCACATTTTAACTGTGACCGCGATTGTTGACCG 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5147 gcaaatcgggttcaacgtaaacccgggaatttcattttggaagagtgatgagctaca 5206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 TCCACACACGACGACTGGGTGGCGCTCCACGTCGAGCGCCCTTATTACTTCGACCCA 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5207 ccaaacgccaagctttagtgtgtgagcccaagaagttgycagtcgaatgtaata 5266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
960 TCGAAGAGACACGCTGCGGTGTTTGCAGGTCTTATGGCTATTATTACCG--CAGATTA 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5267 tgaagcagaagagtcgtggtttatattggttcgcaatcggttgatltcagggaaatga 5326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1017 TAGTGGGGAATTTGTTGGTGGCATCCGTTCCATGATGTGAATCAAAACGACGACGAGC 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5327 tagtggcgagttcgtcgtgtgccttgctcgaagtgtccttgctgctcagca-----ag 5380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1077 AACTTTATTGCTGCTGCGCGGATTTGGCTGACTCCGATGCCACACAGAAATATGAGA 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5381 cgttctttagcaggtgtgtgtattgttcggaatcccaagcagagctagaagaaga 5440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1137 AACTGGTGAATTTGAACCCATG 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5441 aacgaaaaataaattccaacgatg 5465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4  
223756  
ID 223756 standard; DNA: 1659 BP.  
XX  
AC 223756;  
XX

14-JAN-2000 (first entry)  
E. coli entc DNA.  
Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
pathogen inducible promoter; antipathogenic protein; toxin;  
antifungal protein; albumin-type protein; hypersensitive response;  
entc; ss.  
Escherichia coli.

Key Location/Qualifiers  
CDS 360..1535  
FT /tag= a  
FT /product= "entc"

MO950423-A2.

07-OCT-1999.

25-MAR-1999; 99MO-EP02176.

31-MAR-1998; 98US-0080203.

03-APR-1998; 98US-0080625.

(MOGE-) MOGEN INT NV.  
(UYLE-) RIJSDUNIV LEIDEN.  
(UYNT-) UNIV NIJMEGEN.

Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH,  
Van Tegelen LJP, Willems GJ, Croes AF, Stulver MH, Custers J;  
Simons LH, Melchers LS, Bol JF;

WPI: 1999-610856/52.  
P-PSDB: Y50335, Y33698

Method for inducing pathogen resistance in plants

Claim 5; Page 51-53; 66pp; English.

This invention describes a novel method for the induction of pathogen resistance in plants, by transformation with an expression cassette harboring a gene coding for an isochorismate synthase (ICS). A pathogen inducible promoter can be used to drive expression of a heterologous protein. The heterologous protein used in the method of the invention is an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins, saccharide oxidase, oxalate oxidase, magainins, toxins from *Bacillus thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*, *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*, *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal gliadin and wheat-alpha-amylase, or a protein that can induce a hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and CC N-protein from tobacco. This sequence encodes the *Escherichia coli* isochorismate synthase entc gene which is described in the method of the invention.

Sequence 1659 BP; 379 A; 443 C; 454 G; 383 T; 0 other;

Query Match 6.3%; Score 76; DB 20; Length 1659;  
Best Local Similarity 46.6%; Pred. No. 4.8e-15;  
Matches 284; Conservative 0; Mismatches 320; Indels 6; Gaps 1;

563 ATGATGAATGTTATTATTCAGCAACACCGAAGCGTTAGTGGCATGTCAAGTGTGAG 622

939 atggtggtcctcgtctggtgggccaagccggaactgcgtcaagtaagaagcgtgagctt 998

623 TCCGTACGCGCGGCGGTGGGACAAAGCCGCGGAGCGATGCGTACGATATTCG 682

999 ttagctccattcgttagccggttcgcgcgtgtcgcagcgatgaagatgctcgtatcg 1058

683 CGTTAGCGAAGCGTTGTTAGCCAGTCAGAAAAACCGCATTTGAACATCATATGTCGTGG 742



```
Db 1059 aagcaggtatcgtctgctggcgctcagaaaaagatcgcatacgtacgtacgtgtaacc 1118
Qy 743 CAAGATATCGACACCGCTTGCAAGACGTGACGACGTGCTAAAGTCCGCCATCCAA 802
Db 1119 agcgatgtaagagtagtgcgcgaacgtagtgaattacacgttccttcctccac 1178
Qy 803 GTTACTCAAAAATAGCAAGTTCAGCATTTGTACACCAATATACAGGGACATTCGG 862
Db 1179 agctgataccacgcgcgcgcgtctggtcatctgcgaactcccttgaagtaagcgaatt 1238
Qy 863 CACATTTAGTGTGACCCGCGATTTGTGACCGCTTGCAATCCAAACGACGACTGGTGCG 922
Db 1239 cgcgaagaagaacgcactgactctgctctgctgcatccgaccccgctgagcgtgt 1298
Qy 923 TCCACGTAAGAGCGCCCTGATTTACATTTGCGACCCATGAGAAGACACCTCGTGCTGT 982
Db 1299 tcccgatcagcgccgacccaggtattcgtgaactggaacccgttcgacccgcgaactgt 1358
Qy 983 TTGCAAGTCTATTTGCTATTTTACCGCAGATATAGTGGGAATTTGTGTGGCATCC 1042
Db 1359 ttggcgcatctgtggttctgtgtagcagcgaagtaacgcggaattggtgtagcatcc 1418
Qy 1043 GTTCCATGATGTGAATCAACGACGACGACGACCACTTATTTGCTGGTGGCGGATTC 1102
Db 1419 gctcgcggaagctcgcggaataatcagtgcg-----tctgttgcgcgagcggtatgt 1472
Qy 1103 TGGCTGACTCCGATCGCAACCAAGATATGAGAACTGGGTTGAATTTGAACCCATGC 1162
Db 1473 tgcctgcgtcgaacccgttggtgtagtgcgcgaacagcgccaacttctacatgt 1532
Qy 1163 GGCATTTGTT 1172
Db 1533 tgaacgttlt 1542
```

```
RESULT 5
ID 223758 standard; DNA; 2078 BP.
AC 223758;
XX
XX 14-JAN-2000 (first entry)
DE C. roseus ICS DNA.
XX
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KW pathogen inducible promoter; antipathogenic protein; toxin;
KW antifungal protein; albumin-type protein; hypersensitive response;
KW isochorismate pyruvate synthase; ss.
XX
XX Catharanthus roseus.
OS
XX
XX Key Location/Qualifiers
XX CDS 31..1773
XX FT /tag= "a"
XX FT /product= "ICS"
XX
XX WO9950423-A2.
XX
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
XX PR 03-APR-1998; 98US-0080625.
XX
XX (MOSE-) MOSEN INT NV.
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (UYNI-) UNIV NIJMEGEN.
XX
XX Linhorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
XX Van Tegelen LJP, Willems GJ, Croes AF, Stuijver MH, Custers J;
```

```
PI Simons LH, Melchers LS, Bol JF;
XX
XX WPI, 1999-610856/52.
DR P-PSDB; Y50340.
XX
XX Method for inducing pathogen resistance in plants
XX
XX Claim 5; Page 60-62; 66pp; English.
XX
XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magalins, toxins from Bacillus
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
CC Catharanthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Citrus,
CC Lathyrus, Clitoria, Allium seeds, Aralia and Impatiens and albumin-type
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat-alpha-amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and
CC N-protein from tobacco. This sequence encodes the Catharanthus roseus
CC isochorismate synthase ICS protein which is described in the method of
CC the invention.
XX
XX Sequence 2078 BP; 596 A; 423 C; 417 G; 642 T; 0 other;
```

```
Query Match 4.8%; Score 57; DB 20; Length 2078;
Best Local Similarity 50.1%; Pred. No. 1.2e-08;
Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;
```

```
Qy 806 TACTCAAAAATAGCAAGTTCAGCATTTGTACACCAATATACAGGGACATTCGCGCAC 865
Db 1385 taagaagaagcttaagaagaattcaacatcttatgtctgattgcaagcagatcagctctg 1444
Qy 866 ATTTAAGTGTGACCGGATTTGTGACCGCTTGCAATCCAAACGACGACTGGTGCGCTCC 925
Db 1445 aagatgatgatgtcaagaatttcttcctcctcctcctcctcctcctcctcctcctcctc 1504
Qy 926 CACGTGAGCGGCCCTGATTTACATTTGCCAGCCATAGAGAACACCTCGTGCTTGTTC 985
Db 1505 ctatgaagaatgacgcgaatattatctgcggaataatgaatgaatgaatgaatgaatga 1564
Qy 986 CAGGTCTATTTGCTATTTTACCGCAGATATAGTGGGAATTTGTGTGGCATCCGTT 1045
Db 1565 ctggccctgtgttcttccttggaagagctcagagtgatcttctgttggaataagatcgt 1624
Qy 1046 CCATGTATGTGAATCAACGACGACGACGACCACTTATTTGCTGCTGCCGGGATTTGTGG 1105
Db 1625 cc-----ttgattggaagaagatgcgcggtcattataatatacgaggtcttggttag 1678
Qy 1106 CTGACTCCGATGGCAACCAAGATATGAGAACTGGGTTGAA 1148
Db 1679 aaggaagtgatccagctctagaatgcgaagaaactagaatcgtcaa 1721
```

```
RESULT 6
ID V52324/C
XX V52324 standard; DNA; 10357 BP.
XX
XX V52324;
XX
XX 23-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:191.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
OS
```

PN WO9818931-A2.  
XX 07-MAY-1998.  
PD  
XX 30-OCT-1997; 97NO-US19588.  
PE  
XX 31-OCT-1996; 96US-0029960.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;  
DR WPI: 1998-272225/24.  
XX  
XX Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS  
XX Claim 1; Page 1151-1157; 1409pp; English.  
XX  
XX The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
CC  
XX Sequence 10357 BP; 2979 A; 2388 C; 1821 G; 3169 T; 0 other:  
SQ

Query Match 3.5%; Score 42.2; DB 19; Length 10357;  
Best local similarity 47.5%; Pred. No. 0.0027;  
Matches 162; Conservative 0; Mismatches 173; Indels 6; Gaps 1;

QY 821 AAGTTGACGANTTTACACACCATTTACAGGGGACATTTGCGACATTTAAGTGTGACCG 880  
DB 6152 ACGCATGACATTTGACGACGCTGCTCAAGGGCGTTGCTTCCAGAACTCATCGCAAGG 6093  
QY 881 CGATTTGTGACCGCTTGCACACACGACATGCGGTGCGTCCAGCTGAGAGCGCC 940  
DB 6092 ATGCGTTGAAAGCAACGCTTCTGCTGGAACCTTTCTGAGACACAAAGATTGAGCGA 6033  
QY 941 TGTATTACATTTGCGACCCATGAGAAGACACCTGCTGGCTTTTGCAGCTCTATTGGCT 1000  
DB 6032 TGAACGATATCTATGAACTGGAACGAAACGAGGAGTATAGCGTGGGCAATCGGCT 5973  
QY 1001 ATTTTACCAGATTAATAGTGGCAATTTGTGTTGCGATCCGTTCCATGTATGAAATC 1060  
DB 5972 ACTTGTGTGACGACGCTGATGATGATTTGGCATTCGCCATCGCAACA-----TGAATC 5919  
QY 1061 AAACGACGACGACCAACTTATTGCTGTGCGGGATGTGGTGATCGCATCGCATGCGC 1120  
DB 5918 TCAGAAATTCAGAAAGCTTATGTGACGCTGGGCGAGGATTTACGACATCTATCGCC 5859  
QY 1121 AACAGAAATGAGAAACTGGTTGAATTTGAACCATG 1161

DB 5858 AAAACGATACCAAGAAACCATTAACAAAGCAAAATCTATG 5818

RESULT 7  
ID 223757 standard; DNA; 5057 BP.  
XX  
AC 223757;  
XX  
DT 14-JAN-2000 (first entry)  
XX  
DE P. fluorescens DNA encoding ICS orfA and orfD.  
XX  
KW Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KW pathogen inducible promoter; antipathogenic protein; toxin;  
KW antifungal protein; albumin-type protein; hypersensitive response;  
KW isochorismate pyruvate synthase; ss.  
XX  
OS Pseudomonas fluorescens.  
XX  
FH Key Location/Qualifiers  
FT CDS 207..1382  
FT /tag= a  
FT /product= "isochorismate synthase orfA"  
FT CDS 4516..4851  
FT /tag= b  
FT /product= "isochorismate pyruvate lyase orfD"  
XX  
PN WO9950423-A2.  
XX  
PD 07-OCT-1999.  
XX  
PE 25-MAR-1999; 99WO-EP02176.  
XX  
PR 31-MAR-1998; 98US-0080203.  
PR 03-APR-1998; 98US-0080625.  
XX  
PA (MOGE-) MOGEN INT NV.  
PA (DYLE-) RIJKSUNIV LEIDEN.  
PA (UYNT-) UNIV NIJMEGEN.  
XX  
PI Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;  
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuijver MH, Custers J;  
PI Simons LH, Melchers LS, Bol JF;  
XX  
DR WPI: 1999-610856/52.  
DR P-PSDB: Y50336, Y50337, Y500338, Y50339.  
XX  
PT Method for inducing pathogen resistance in plants  
XX  
PS Claim 5; Page 54-58; 66pp; English.  
XX  
XX This invention describes a novel method for the induction of pathogen  
CC resistance in plants, by transformation with an expression cassette  
CC harboring a gene coding for an isochorismate synthase (ICS), a pathogen  
CC inducible promoter can be used to drive expression of a heterologous  
CC protein. The heterologous protein used in the method of the invention is  
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
CC saccharide oxidase, oxalate oxidase, magalins, toxins from Bacillus  
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,  
CC Amaryanthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Cnicus,  
CC Lathyrus, Clitoria, Allium seeds, Arazilia and Impatiens and albumin-type  
CC proteins, such as chitinase, napin, barley trypsin inhibitor, cereal  
CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and  
CC N-protein from tobacco. This sequence encodes the Pseudomonas fluorescens  
CC isochorismate synthase orfA protein and the isochorismate pyruvate lyase  
CC orfD protein which are described in the method of the invention.  
XX  
SQ Sequence 5057 BP; 1171 A; 1554 C; 1302 G; 1030 T; 0 other;

	Query Match	3.5%	Score 41.6:	DB 20:	Length 5037:
	Best Local Similarity	45.1%	Pred. No. 0.0029:		
	Matches 241:	Conservative 0:	Mismatch 284:	Indels 9:	Gaps 2:
OY	635	CGGTGCTGGGCAAGCCGGCGGGAGAGATGGGCGGTACATATATGCGTTAGGCGAAG	694		
DB	832	cgctggcggggttcagcccaagcgcgatgcgatccagaagccgagccggcgcaatgcgaact	891		
OY	695	CGTTGTAGCCAGTCGAGAAAAACCGCATTTGACATCAATATGTCGTGGCAGTATCAGCA	754		
DB	892	ggttcgtacatcgcgaaaaagatacactacacgacgaggttcgttcgacagagcatcgtca	951		
OY	755	CACGCTTGCAAGACGTGCAGACGCTGCTTAAGAGTCCCGGCAATGCCAAGTTACTCAAAA	814		
DB	952	gccaaacttgggaaactctgacacgcagctgaaatgtgctcgcaagccctccatcaccga	1011		
OY	815	ATTAACCAAGTTCAGATTGTGTACACCAAAATTAACAGGGGACATTCGCG--CACATTTAA	871		
DB	1012	cgccgcgcgtctgcaacctcgcacctcgcacatcgaaagtaacgtgcgcaaccgcgctgat	1071		
OY	872	GTGTGACCGCGATTTGTTGACCGCTTGCAATCCAAACACAGCACTGGGTGGCGTCCACGTC	931		
DB	1072	cggccttcgagctgtgcctgcgcgtctgcaccccaacacgcgctgtgycggtccccaacg	1131		
OY	932	AAGCGCCCTGATTACATTCGCGACCCATGAGAAACACTCTGTCGTTGTTGTCAGGTC	991		
DB	1132	agcgcgcccgcgcgctgcgtctgccttcgltcacacctcgaagcgcgctgtccacccgca	1191		
OY	992	CTATTGGCATTTTACCGGAGATATAGGGGGAATTTGTGGTTGGCATCGTTCATGT	1051		
DB	1192	tgtgtgtgtgtgtgtgcgtatgccagggcaatggcgaaatgggtcgttaacgactcgttcgca	1251		
OY	1052	ATGTGAATCAACAGCAGCGACGAGCACTTATTTGCTGTGTCGCGGATTTGGCTGACT	1111		
DB	1252	cggctcagcgcaaaacgaagtcg-----ccgttcgcgcgagcgcatcgtltgaagcct	1305		
OY	1112	CCGATGCCCAACACAAATATGAGAAACTGGGTTGAATTGTAACCCATCGCGC	1165		
DB	1306	caagcccgactcgcgaatggcgagaagtcacacaaacttgcacacatcgtgc	1359		
RESULT 8					
ID	V21209	standard: DNA: 1664976 BP.			
XX	V21209;				
AC	10-NOV-1998	(first entry)			
XX	DE	Methanococcus jannaschii circular chromosome.			
XX	XX				
KW	Metanococcus jannaschii: methanogenic archaeon; circular chromosome;				
XX	genome; autotrophic; extrachromosomal element; identification; ds.				
OS	Methanococcus jannaschii.				
XX	XX				
PN	W09807830-A2.				
XX	PD	26-FEB-1998.			
XX	PF	22-AUG-1997; 97WO-US14900.			
XX	PR	22-AUG-1996; 96US-0024428.			
XX	PA	(GENO-) INST GENOMIC RES.			
XX	PA	(UNIT) UNIT ILLINOIS FOUND.			
XX	PA	(UYO ) UNIT JOHNS HOPKINS SCHOOL MEDICINE.			
XX	PI	Bult CJ, Smith HO, Venter JC, White OR, Woese CR;			
XX	XX	WPI: 1998-169145/15.			
XX	XX				
PT	Complete genome sequence of methano-genic archaeon, Methanococcus				

[illegible]

AC 296384;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE S. pneumoniae derived DNA from ORF #212.  
XX  
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;  
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN W09806734-A1.  
XX  
PD 19-FEB-1998.  
XX  
PE 15-AUG-1997; 97WO-US14436.  
XX  
PR 16-AUG-1996; 96US-0024022.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
PI Scodola RK;  
DR WPI; 1998-159452/14.  
DR P-PSDB; Y86058, Y86059, Y86060.  
XX  
PT Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity  
XX  
PS Claim 4; Page 242; 640pp; English.  
XX  
CC This invention describes novel isolated Streptococcus pneumoniae  
CC polynucleotides (see 296173-296494) and their encoded proteins (see  
CC Y85792-Y86182). The DNA, vectors and host cells described in the method  
CC of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which  
CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.  
XX  
SQ Sequence 1167 BP; 319 A; 280 C; 228 G; 340 T; 0 other;  
  
Query Match 2.8%; Score 33.8; DB 19; Length 1167;  
Best Local Similarity 62.4%; Pred. No. 0.5;  
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
  
QY 1055 TGAATCAACGACGACGACCAATTATTTGCTGTGCGGATGTGCTGACTCG 1114  
DB 1149 TGAATTCACAAAATCAAAAGACCTATGTGCAAGCTGGGGGAGATGTCTACGACTCTA 1090  
QY 1115 ATGGCAACAAGATATGAGAAAC 1139  
DB 1089 TCGCCCAAAAGCAATACCAAGAAC 1065  
  
RESULT 10  
V74332  
ID V74332 standard; DNA; 3647 BP.  
XX  
AC V74333;  
XX  
DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #21.  
XX

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
XX  
OS Staphylococcus aureus.  
XX  
FH Key location/Qualifiers  
FH misc-feature 301..360  
FT /tag- a  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc-feature 2101..2160  
FT /tag- b  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
XX  
PN EP786519-A2.  
XX  
PD 30-JUL-1997.  
XX  
PF 07-JAN-1997; 97EP-0100117.  
XX  
PR 05-JAN-1996; 96US-0009861.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
PI Rosen CA;  
DR WPI; 1997-374922/35.  
XX  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
PS Claim 1; Page 277-279; 3271pp; English.  
XX  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
XX  
SQ Sequence 3647 BP; 1315 A; 550 C; 619 G; 1037 T; 126 other;  
  
Query Match 2.7%; Score 32.4; DB 18; Length 3647;  
Best Local Similarity 45.3%; Pred. No. 2.8;  
Matches 163; Conservative 0; Mismatches 191; Indels 15; Gaps 1;  
  
QY 811 AAAAATAGCAAGTTGACATTTGTACACACCAATTACAGGGGACATGCGGCAATTGA 870  
DB 563 aaatagcaagttatgcatgtaagtgaaagtcacaaagtgtaaaataacataaattta 622  
QY 871 AGTGTGACCGGCAATGTTGACCGCTTGATCCACACGACGACTGTGGGTCCACAGT 930

Db 623 tcgccaatgacagttattgcgaattattaccacagcgttcacagtgccacaaa 682  
QY 931 GAAGGGCCCTGATTACATTCGACCATGAGAGACACCTCGGCTGTTTGCAGCT 990  
Db 683 ttacgtgaatgaaagaatataatgaacaatataccacaataacggcgcttataagtg 742  
QY 991 CCTATTGCTATTATTACGCGAGATAGTGGGAATTTGTGGTGGCATCCGTTCCATG 1050  
Db 743 ggtgttgataataatgttaataactagatttgcaatcgaatcgaacgagtg 802  
QY 1051 TATGTGAATCAACGACGAGACCACTTATTGCTGTCGCGGATTTGGCTGAC 1110  
Db 803 atgataagat-----gagcagatatacaacgtagaagcgtgtgtgtatataatg 856  
QY 1111 TCCGATGGCGAACAAGATATGAGAACTGGGTTGAAATTTGAACCCATGCGCAATG 1170  
Db 857 tctattcctgaaagaactgaatgaacgaatgaaagcctaagaagcttattgagtg 916

## RESULT 11

X13035 ID X13035 standard; DNA; 16870 BP.

XX AC X13035;

XX DT 19-MAR-1999 (first entry)

XX DE Enterococcus faecalis genome contig SEQ ID NO:98.

KW Enterococcus faecalis; contig; detection: Enterococcal infection; vaccine; attenuation; computer readable medium; ds.

XX OS Enterococcus faecalis.

XX PN W09850555-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08985.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Dillon PJ, Kunsch CA;

XX DR WPI: 1999-045171/04.

XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides - used to develop products for the detection of Enterococcus and for use in vaccines for prevention or attenuation of Enterococcus infection.

XX PS Claim 1; Page 634-642; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it  
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
XX X12938 to X13919 represent these nucleotide sequences which are primary  
XX nucleotide sequences, also known as contigs. The computer-based system  
XX can identify fragments of the Enterococcus faecalis genome with  
XX commercial importance. The products can be used to detect the presence  
XX of Enterococcus faecalis in samples. They can also be used for  
XX diagnosing Enterococcal infection in an animal and monitoring  
XX progression of disease, and for identifying agents which can be used to  
XX modulate the growth or pathogenicity of Enterococcus faecalis, or  
XX another related organism, in vivo or in vitro. In particular the  
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
XX can be used in vaccines to prevent or attenuate an Enterococcal  
XX infection.

SQ Sequence 16870 BP; 5332 A; 3410 C; 2947 G; 5177 T; 4 other;

Query Match 2.6%; Score 31.6; DB 20; Length 16870;

Best Local Similarity 52.2%; Pred. No. 13; Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 461 GTCGCAACAGACCCCTGCAGTTATCCGACGTTACGACTGGCACAAATATTCTGCGCT 520  
Db 7311 gtcgtaattgactcgtggaatcgtcccaatcgcaacgacatcataaagtagtgcatt 7370  
QY 521 TAGCTGACGACGCGAATACGTATCATGTGTTTAAAGCAGACATGTAATGTTATT 580  
Db 7371 gatttccccacatctaaagcactgtttttcttaaatcaattgtgatcgcttctc 7430  
QY 581 CAGCAACACCGGAA 594  
Db 7431 aggtccatcgta 7444

## RESULT 12

Q29937 ID Q29937 standard; cDNA; 2032 BP.

XX AC Q29937;

XX DT 11-MAR-1993 (first entry)

XX DE PEP4 gene encoding Pichia pastoris proteinase A.

KW Proteolytic; vacuolar; aspartyl protease; EGF; GRF; IGF-1; ss.

XX OS Pichia pastoris.

XX FH Key Location/Qualifiers

XX FT CDS 239..1468

XX FT /\*tag= a

XX PN W09217595-A.

XX PD 15-OCT-1992.

XX PF 31-MAR-1992; 92MO-US02521.

XX PR 01-APR-1991; 91US-0678916.

XX PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.

XX PI Gleeson MA, Howard BD;

XX DR WPI: 1992-366264/44.

XX PT P-PSDB; R28030.

XX PT New isolated DNA fragment from Pichia comprising gene influencing proteolytic activity - used to generate proteolytic deficient strains, as hosts for expression of proteolytically sensitive proteins

XX PS Claim 7; Page 67-69; 108pp; English.

XX The Pichia pastoris PEP4 gene encoding proteinase A was identified in  
XX a lambda based EMBL3 P. pastoris genomic DNA library by its ability  
XX to hybridise with a radiolabelled fragment of the homologous  
XX Saccharomyces cerevisiae PEP4 gene. The P. pastoris PEP4 gene was  
XX cloned by isolating positive plaques contg. the hybridising  
XX recombinant phage DNA. The PEP4 prod., a vacuolar, aspartyl  
XX protease capable of self-activation, as well as subsequent  
XX activation of additional vacuolar proteases, e.g. carboxypeptidase Y  
XX and proteinase B may be mutated to create a proteolysis-deficient  
XX strain of Pichia which may be used for recombinant expression of  
XX proteolytically sensitive proteins, e.g. EGF, GRF and IGF-1.  
XX See also Q29938-41.



CC of transgenic plants.  
XX  
SQ Sequence 2161 BP; 603 A; 420 C; 530 G; 608 T; 0 other;  
  
Query Match  
Best Local Similarity 49.7%; Pred. No. 8.3;  
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
  
QY 572 TGTATTTCAGCACACCGGAACGGTTAGTGCCTATGTCAGGTGTCAGATGCTACGG 631  
DB 1144 tttaagtcgacgcagccagaatttgacacgtggaagaagaagattgtaac 1203  
QY 632 CGGCGGTGCTGGGACAAAGCGGCGCGGACGATGCGCTGACGATGCGCTTAGCG 691  
DB 1204 gaccactggtggaagaagaagaagacacctgattgagatgattgtgtaa 1263  
QY 692 AAGCGTTGTTAGCCAGTCAAGAAAACCGCATTTGACA 728  
DB 1264 tgcagatgttaaaagatgagaacaacgcgcagagca 1300  
  
RESULT 15  
V52138/c  
ID V52138 standard; DNA; 13926 BP.  
XX  
AC V52138;  
XX  
DT 23-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:5.  
XX  
KM Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN MO9818931-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97MO-US19588.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;  
XX  
DR WPI; 1998-272225/24.  
XX  
PS  
XX  
CC Claim 1; Page 174-182; 1409pp; English.  
CC  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide

CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
XX  
SQ Sequence 13926 BP; 3889 A; 2986 C; 2460 G; 4591 T; 0 other;  
XX  
Query Match  
Best Local Similarity 45.7%; Pred. No. 28;  
Matches 106; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
  
QY 511 ATTGTCGTTAGCTGACGAGCGCAATACATATGCTTTTAAGCGCATGATGAA 570  
DB 1842 AATCGTATGTTGTTAGAGAGAGAGCGGCTACATGCTATGTGGAACATGACGATG 1783  
QY 571 TTGTTATTTCAGCACACCGGAACGGTTAGTGCCTATGTCAGGTGCTCAGATCGTACG 630  
DB 1782 ACAGTGATTTCCATGAGCCCGAGACCTCTTTTGTGCAAAATGATCCGAGTTGACAA 1723  
QY 631 GCGGCGGTGCTGGGACAAAGCGGCGGACGATGCGCTGACGATTCGCTTAGCG 690  
DB 1722 CGACCAATGAAAGGAGGAGACTCAGCGTGGGTAACTACCAAGAAATCTTGAACAGGCC 1663  
QY 691 GAAGCGTTGTTAGCCAGTCAAGAAAACCGCATTTGACATCAATATGTCGTGG 742  
DB 1662 AGTTGTTGGAACAGATCCCAAAATTCGCTGTAATAATGATGATGTGG 1611

Search completed: November 25, 2000, 02:08:35  
Job time: 3907 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 20:03:37 : Search time 85.07 Seconds  
(without alignments)  
2128.083 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197

Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTCGATGA 1197

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/PCUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1197	100.0	1197	4	US-08-774-104A-1
2	33.8	2.8	7218	1	US-08-232-463-14
3	30.8	2.6	2032	1	US-08-088-633-1
4	30.8	2.6	2032	1	US-08-245-756-1
5	30.8	2.6	2032	2	US-08-441-750-1
6	30.8	2.6	2032	3	US-08-441-751-1
7	30.8	2.6	2032	6	PCR-US92-02521-1
8	30.8	2.6	3640	4	US-08-627-873-6
9	30.6	2.6	2161	4	US-09-001-825-4
10	30.6	2.6	2161	4	US-09-001-826-25
11	30.4	2.5	458	5	US-09-141-000-4
12	29.4	2.5	1419	2	US-08-242-098-39
13	29.4	2.5	1650	4	US-09-001-826-24
14	29.4	2.5	2380	2	US-08-572-951-3
15	29.4	2.5	24417	4	US-08-846-762-1
16	29.2	2.4	554	4	US-09-121-179-2
17	29	2.4	1119	1	US-07-854-596B-30
18	29	2.4	1257	1	US-07-854-596B-25
19	29	2.4	1317	1	US-07-854-596B-18
20	29	2.4	1335	1	US-07-854-596B-14
21	29	2.4	1458	1	US-07-854-596B-42
22	29	2.4	1466	4	US-08-749-902-2
23	29	2.4	1467	1	US-07-854-596B-46
24	29	2.4	1512	1	US-07-854-596B-27
25	29	2.4	2253	1	US-07-854-596B-39
26	29	2.4	2589	1	US-07-854-596B-34

27	28.8	2.4	1908	5	US-09-041-991A-3	Sequence 3, Appl
28	28.6	2.4	1079	5	US-08-454-196-16	Sequence 16, Appl
29	28.6	2.4	1079	5	US-09-064-033-16	Sequence 16, Appl
30	28.6	2.4	3164	1	US-07-723-002C-3	Sequence 3, Appl
31	28.6	2.4	5904	6	US-08-309-512-1	Sequence 1, Appl
32	28.6	2.4	5904	6	PCR-US92-08756A-1	Sequence 1, Appl
33	28.4	2.4	2524	1	US-08-317-522A-1	Sequence 1, Appl
34	28.4	2.4	2524	1	US-08-439-818A-1	Sequence 1, Appl
35	28.4	2.4	2524	3	US-08-751-965-1	Sequence 1, Appl
36	28.4	2.4	2524	3	US-08-738-975-1	Sequence 1, Appl
37	28.4	2.4	2524	4	US-08-728-626-1	Sequence 1, Appl
38	28.4	2.4	2524	5	US-08-808-599A-1	Sequence 1, Appl
39	28.4	2.4	3477	1	US-08-141-324-13	Sequence 13, Appl
40	28.4	2.4	3477	2	US-08-541-902-13	Sequence 13, Appl
41	28.4	2.4	6241	2	US-08-570-311-25	Sequence 25, Appl
42	28.2	2.4	4406	1	US-08-369-043-1	Sequence 12, Appl
43	28.2	2.4	43676	5	US-09-356-952-12	Sequence 12, Appl
44	28	2.3	891	7	5164490-2	Patent No. 5164490
45	27.8	2.3	685	5	US-08-792-832A-47	Sequence 47, Appl

#### ALIGNMENTS

RESULT 1

US-08-774-104A-1

Sequence 1, Application US/08774104A

Patent No. 5919687

GENERAL INFORMATION:

APPLICANT: Chatterjee, Subroto

TITLE OF INVENTION: RECOMBINANT N-SMASES AND NUCLEIC ACIDS

TITLE OF INVENTION: ENCODING SAME

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dikey, Bronstein, Roberts & Cushman, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/774,104A

FILING DATE: 12/24/96

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Corless, Peter F.

REGISTRATION NUMBER: 33,860

REFERENCE/DOCKET NUMBER: 46906

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1197 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE:

ORIGINAL SOURCE:

US-08-774-104A-1





Query Match 2.6%; Score 30.8; DB 1; Length 2032;  
Best Local Similarity 51.4%; Pred. No. 2.1;  
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 444 AGCCAAAGTGTGTTTGGTGGCAACAGACCCTGCACTTATCCGACACGTTACGACTGGC 503  
DB 307 AGCCAAAGTGTGTTTGGTGGCAACAGACCCTGCACTTATCCGACACGTTACGACTGGC 366  
QY 504 ACAATATTATTCGTGCTAGCTGAGAGCGCAATACGTATCATGTGTTTAAAGCGACA 563  
DB 367 CAATTTGGGCAAGTATCTCTGCTGGAACATTAATATGTTTCTGTTCACAGCACA 426  
QY 564 TGATGAATGTTTATTTTC 581  
DB 427 AATGCTTTGTCCAAAGTC 444

RESULT 5  
US-08-441-750-1  
; Sequence 1, Application US/08441750  
; Patent No. 5691166  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHTA PROTEOLYTIC  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,750  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/245,756  
; FILING DATE: 16-May-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/088,633  
; FILING DATE: 06-JULY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 9763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2032 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 239..1468  
; FEATURE:

NAME/KEY: mat-peptide  
LOCATION: 239..1468  
US-08-441-750-1

Query Match 2.6%; Score 30.8; DB 2; Length 2032;  
Best Local Similarity 51.4%; Pred. No. 2.1;  
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 444 AGCCAAAGTGTGTTTGGTGGCAACAGACCCTGCACTTATCCGACACGTTACGACTGGC 503  
DB 307 AGCCAAAGTGTGTTTGGTGGCAACAGACCCTGCACTTATCCGACACGTTACGACTGGC 366  
QY 504 ACAATATTATTCGTGCTAGCTGAGAGCGCAATACGTATCATGTGTTTAAAGCGACA 563  
DB 367 CAATTTGGGCAAGTATCTCTGCTGGAACATTAATATGTTTCTGTTCACAGCACA 426  
QY 564 TGATGAATGTTTATTTTC 581  
DB 427 AATGCTTTGTCCAAAGTC 444

RESULT 6  
US-08-441-751-1  
; Sequence 1, Application US/08441751  
; Patent No. 5831053  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHTA PROTEOLYTIC  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,751  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/245,756  
; FILING DATE: 16-May-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/088,633  
; FILING DATE: 06-JULY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 9763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2032 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 239..1468
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 239..1468
;
US-08-441-751-1

```

Query Match	2.68;	Score 30.8;	DB 3;	Length 2032;
Best Local Similarity	51.48;	Pred. No. 2.1;		
Matches 71; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0

QY	444	AGCCAAAGTCCTTTTGGTCGGCAACAGACCTTCAGATTTTCGACACAGTTAGACACTGGC	503
Db	307	AGCCAAAGTCCTTTGGTCGGCAACAGACCTTCAGATTTTCGACACAGTTAGACACTGGC	366
QY	504	ACAAATTTATTCGTCGTTAGTCGACAGGGCAATACGTATACATGCGTTTAAAGGCACA	563
Db	367	CAATTTTGGCAGATATGTCCTGCTCTGGACATTAATATGTTCTCTGTTCAACGACACA	426
QY	564	TGATGCAATTTGTTTATTC	581
Db	427	AAATGCTTTGTCAGTC	444

RESULT 7  
PCT-US92-02521-1  
; Sequence 1, Application PC/TUS9202521

LOCATION: 239..1468  
PCT-US92-02521-1

Query Match	2.68;	Score 30.8;	DB 6;	length 2032;
Best Local Similarity	51.48;	Pred. No. 2.1;		
Matches 71; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0;

[illegible]

RESULT 8  
US-08-627-873-6  
; Sequence 6, Application US/08627873

NAME/KEY: CDS (exon)  
LOCATION: 886..1402  
NAME/KEY: Intron  
LOCATION: 1403..1478  
NAME/KEY: CDS (exon)  
LOCATION: 1479..3637  
US-08-627-873-6

Query Match 2.6%; Score 30.8; DB 4; Length 3640;  
Best Local Similarity 61.0%; Pred. No. 3;  
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 65 TTGAACAACGATTTGGCGCTTACTTTCGACACCCGCTGATGATCTTACGTTTG 124  
DB 2394 TTGACATACCTCAAGTCGAGCTACCACTGTTCAGCCGCTGATGATGACAGTGTG 2453  
OY 125 GCATTGGCGCAATCGCTACGCG 146  
DB 2454 TTGTGAGGTGATCCCTCGAC 2475

RESULT 9  
US-09-001-826-4  
Sequence 4, Application US/09001826A  
Patent No. 5965727  
GENERAL INFORMATION:  
APPLICANT: SONG, HEE-SOOK  
APPLICANT: BROTHERTON, JEFFREY E.  
APPLICANT: WIDHOLM, JACK M.  
TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
FILE REFERENCE: UI001.C1  
CURRENT APPLICATION NUMBER: US/09/001,826A  
EARLIER FILING DATE: 1997-12-31  
EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
EARLIER FILING DATE: 1997-07-25; 1996-07-26  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
SEQ ID NO 4  
LENGTH: 2161  
TYPE: DNA (cDNA)  
ORGANISM: Nicotiana tabacum  
US-09-001-826-4

Query Match 2.6%; Score 30.6; DB 4; Length 2161;  
Best Local Similarity 49.7%; Pred. No. 2.5;  
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 572 TGTATTTCACGACACCGAAGCGTTAGTGGCTATGTCAGTGGTACATCGCTACGG 631  
DB 1144 ttttagttgcacgcgaccagaattttgacacgtgtgaagaagaataattgttaac 1203  
OY 632 CGGCGCTGCTGGACAAAGCCGCGCGGACGATGGCGCTGACGATATCGGTTAGGCG 691  
DB 1204 gaccactgctggaagaagaagaagaagaacacccatgatgagatgtgtgtgga 1263  
OY 692 AAGCCTTTGACCGACGAGAAAACCGCATTTGAACA 728  
DB 1264 tgcagatgtltaagaagatgagaacaacgcgcagagca 1300

RESULT 10  
US-09-001-826-25  
Sequence 25, Application US/09001826A  
Patent No. 5965727  
GENERAL INFORMATION:  
APPLICANT: SONG, HEE-SOOK  
APPLICANT: BROTHERTON, JEFFREY E.  
APPLICANT: WIDHOLM, JACK M.  
TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
FILE REFERENCE: TRANSFORMATION

FILE REFERENCE: UI001.C1  
CURRENT APPLICATION NUMBER: US/09/001,826A  
CURRENT FILING DATE: 1997-12-31  
EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
EARLIER FILING DATE: 1997-07-25; 1996-07-26  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
SEQ ID NO 25  
LENGTH: 2161  
TYPE: DNA (cDNA)  
ORGANISM: Nicotiana tabacum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: LOCATION: 90..1940  
US-09-001-826-25

Query Match 2.6%; Score 30.6; DB 4; Length 2161;  
Best Local Similarity 49.7%; Pred. No. 2.5;  
Matches 78; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 572 TGTATTTCACGACACCGGAAAGCGTTAGTGGCTATGTCAGTGGTACATCGCTACGG 631  
DB 1144 ttttagttgcacgcgaccagaattttgacacgtgtgaagaagaataattgttaac 1203  
OY 632 CGGCGCTGCTGGACAAAGCCGCGCGGACGATGGCGCTGACGATATCGGTTAGGCG 691  
DB 1204 gaccactgctggaagaagaagaagaagaacacccatgatgagatgtgtgtgga 1263  
OY 692 AAGCCTTTGACCGACGAGAAAACCGCATTTGAACA 728  
DB 1264 tgcagatgtltaagaagatgagaacaacgcgcagagca 1300

RESULT 11  
US-09-141-000-4/c  
Sequence 4, Application US/09141000  
Patent No. 6054295  
GENERAL INFORMATION:  
APPLICANT: Chen, Pang  
TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR  
FILE REFERENCE: 199999Y  
CURRENT APPLICATION NUMBER: US/09/141,000  
CURRENT FILING DATE: 1998-08-26  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 458  
TYPE: PRT  
ORGANISM: Human  
US-09-141-000-4

Query Match 2.5%; Score 30.4; DB 5; Length 458;  
Best Local Similarity 9.1%; Pred. No. 1.1;  
Matches 29; Conservative 95; Mismatches 199; Indels 0; Gaps 0;

OY 664 GATGCGCTGACGATATCGCTTAGCGGCAAGCGTTGTTAGCCATGACAGAAAACCGCAT 723  
DB 458 BMT..K...MDK..BMC..M.NR.D.BTWASA.Y...AK.KMCTYY.H.KD.CT.RH.. 399  
OY 724 GAACATCAATATGTCGTGGCAAGTATCACGACACGCTTGCAAGAGCTGACGAGTGGCTA 783  
DB 398 T.D.BH..M.BT.BH..DKSHSNT.T.TM.AB...M..MKSMMMB...TNN.H..CT.. 339  
OY 784 AAGTGGCGGCGCATCCCAAGTTTACCAAAAATAGACAAGCATTTGATTTGACACACA 843  
DB 338 MS.H.HK.RHRTB..H..S.SYRBC...KMTS...SK.HT.S.AS.C..DWTWC.. 279  
OY 844 ATTACAGGAGACATTTGCGGACACATTTAAGTGTGACCGGATTTGACCGCTTGATCA 903  
DB 278 BB..YHT.HG..AA.TM.HSH.BA.H.KTR.M...TB..DSB.MNR.MMT..B...N.R.S 219

QY 904 ACACGACACTGGGTGGCCCTCCACGTGAAGCGCCCTGATTACATTGGACCATGAG 963  
DB 218 N.YH.YYMR.YCCYBYH.YBC.M.KCBM.GMK.YGT.GSMYYMA..G.NAT.GSR..N 159  
QY 964 AAGACACCTCGTGGCTT 980  
DB 158 C.AYM.TMGT.GSTB 142

RESULT 12  
US-08-242-098-39  
; Sequence 39, Application US/08242098  
; Patent No. 5691185  
; GENERAL INFORMATION:  
; APPLICANT: DICKELY, Francoise  
; APPLICANT: JOHANSEN, Eric  
; APPLICANT: NILSSON, Dan  
; APPLICANT: HANSEN, Egon  
; APPLICANT: STROMAN, Per  
; TITLE OF INVENTION: Lactic Acid Bacterial Suppressor Mutants  
; TITLE OF INVENTION: and Their Use as Selective Markers and as Means of  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/242, 098  
; FILING DATE: 13-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/133,390  
; FILING DATE: 08-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30307/141/PLVI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1419 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-242-098-39

Query Match 2.5%; Score 29.4; DB 2; Length 1419;  
Best Local Similarity 51.1%; Pred. No. 5.1;  
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 68 AACACACTGAATTTGGCGCTTACTTTCGACACCCGCTGATGATCTTACGTTTGCA 127  
DB 584 ATATCACAGATATATCTCATTTCTCAGTAAATATGCTGCAAGTGTCTTACAGCT 643  
QY 128 TTGGCCGCAATCGCTACGGCAAAAACGCGCTACAGCATTAAGAAGTGGCGTTGTTGGTG 187  
DB 644 TTAAAGCAGCTCTTGCAGAAAATATGTTCCAGATCAGCAACTCAGTATTTCAAGTGT 703  
QY 188 CGCAGTCATTTGATG 202

DB 704 CTAAAGATTAGAG 718

RESULT 13  
US-09-001-826-24  
; Sequence 24, Application US/09001826A  
; Patent No. 5965727  
; GENERAL INFORMATION:  
; APPLICANT: SONG, HEE-SOOK  
; APPLICANT: BROTHERTON, JEFFREY E.  
; APPLICANT: WIDHOLM, JACK M.  
; TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE  
; TITLE OF INVENTION: TRANSFORMATION  
; FILE REFERENCE: U1001.C1  
; CURRENT APPLICATION NUMBER: US/09/001,826A  
; CURRENT FILING DATE: 1997-12-31  
; EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140  
; EARLIER FILING DATE: 1997-07-25; 1996-07-26  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text  
; SEQ ID NO 24  
; LENGTH: 1650  
; TYPE: DNA (cDNA)  
; ORGANISM: Nicotiana tabacum  
; US-09-001-826-24

Query Match 2.5%; Score 29.4; DB 4; Length 1650;  
Best Local Similarity 46.4%; Pred. No. 5.6;  
Matches 96; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 522 AGCTGACGACGCGAATACGATCATGTGCTTTAAACGACATGATGATTTTATTTTC 581  
DB 742 agtgtgataaccgaagtcgatatgttgcaagcagaagatgcatcttgtagc 801  
QY 582 AGCACACGCGAAGCGTTAGTGGCTATGTCAGGTGCTCAGATGCTACGGCGGCGTGC 641  
DB 802 atcaagtcagaatctccaccaagaagcaagaagatagtgatcgtgccaatgagc 861  
QY 642 TGGGACACAGCCGCGCGACGATGGCGCTGACGATATGCGCTTAGCGCAAGCGTTTGT 701  
DB 862 aggaaccagaagaagagaggaagatgagtgtagaagataagataagaaactgct 921  
QY 702 AGCCAGTCAGAAAAACCGCATTTGACA 728  
DB 922 aggaatgaaaaagaaagtgtgagca 948

RESULT 14  
US-08-572-951-3  
; Sequence 3, Application US/08572951  
; Patent No. 5824790  
; GENERAL INFORMATION:  
; APPLICANT: KEELING, PETER L.  
; APPLICANT: KNIGHT, MARY E.  
; APPLICANT: GUAN, HANPING  
; TITLE OF INVENTION: MODIFICATION OF STARCH  
; TITLE OF INVENTION: SYNTHESIS IN PLANTS  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; ADDRESSEE: Intellectual Property Group of  
; ADDRESSEE: Pillsbury Madison & Sutro LLP  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

1      CURRENT APPLICATION DATA:
2      APPLICATION NUMBER:   US/08/572,951
3      FILING DATE: 15-DEC-1995
4      CLASSIFICATION: 800
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 08/346,602
7      FILING DATE: 29-NOV-1994
8      CLASSIFICATION: 800
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: 08/263,921
11     FILING DATE: 21-JUN-1994
12     CLASSIFICATION: 800
13     ATTORNEY/AGENT INFORMATION:
14     NAME: Paul N. Kokulis
15     REGISTRATION NUMBER: 16,773
16     REFERENCE/DOCKET NUMBER: 222957/1..02.15C
17     TELECOMMUNICATION INFORMATION:
18     TELEPHONE: (202) 861-3000
19     TELEFAX: (202) 822-0944
20     INFORMATION FOR SEQ ID NO: 3:
21     SEQUENCE CHARACTERISTICS:
22     LENGTH: 2380 base pairs
23     TYPE: nucleic acid
24     STRANDEDNESS: both
25     TOPOLOGY: linear
26     MOLECULE TYPE: CDNA
27     US-08-572-951-3

```

Query Match	2.58	Score	29.4	DB	2	Length	2380
Best Local Similarity	30.18	Pred	No. 7				
Matches	44	Conservative	27	Mismatches	75	Indels	0
						Gaps	0

QY	386	TGGATTTGGATGCAAGCAGCAGACAAATTTGATATACCTTACCCATTCGATCAAACTTAC	445
	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		
	11		
	12		
	13		
	14		
	15		
	16		
	17		
	18		
	19		
	20		
	21		
	22		
	23		
	24		
	25		
	26		
	27		
	28		
	29		
	30		
	31		
	32		
	33		
	34		
	35		
	36		
	37		
	38		
	39		
	40		
	41		
	42		
	43		
	44		
	45		
	46		
	47		
	48		
	49		
	50		
	51		
	52		
	53		
	54		
	55		
	56		
	57		
	58		
	59		
	60		
	61		
	62		
	63		
	64		
	65		
	66		
	67		
	68		
	69		
	70		
	71		
	72		
	73		
	74		
	75		
	76		
	77		
	78		
	79		
	80		
	81		
	82		
	83		
	84		
	85		
	86		
	87		
	88		
	89		
	90		
	91		
	92		
	93		
	94		
	95		
	96		
	97		
	98		
	99		
	100		
	101		
	102		
	103		
	104		
	105		
	106		
	107		
	108		
	109		
	110		
	111		
	112		
	113		
	114		
	115		
	116		
	117		
	118		
	119		
	120		
	121		
	122		
	123		
	124		
	125		
	126		
	127		
	128		
	129		

```

RESULT 15
US-08-846-762-1
: Sequence 1, Application US/08846762A
: Patent No. 5994072
: GENERAL INFORMATION:
: APPLICANT: Lam, Joseph S.
: APPLICANT: Burrows, Lori
: APPLICANT: Charter, Deborah
: APPLICANT: de Kievit, Teresa
: TITLE OF INVENTION: No. 5994072e1 Proteins Involved in the Synthesis and Assembly
: TITLE OF INVENTION: of O-antigen in Pseudomonas Aeruginosa
: FILE REFERENCE: 6580-089
: CURRENT APPLICATION NUMBER: US/08/846,762A
: CURRENT FILING DATE: 1997-04-30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO. 1
: LENGTH: 24417
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-08-846-762-1

```

Query Match	2.5%	Score 29.4	DB 4	Length 24417
Best Local Similarity	51.1%	Pred. No. 30		
Matches 69	Conservative 0	Mismatches 66	Indels 0	Gaps 0

[illegible]

Search completed: November 25, 2000, 01:48:46  
Job time: 20709 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 24, 2000, 18:38:06 ; Search time 852.26 Seconds  
(without alignments) 8683.758 Million cell updates/sec

Title: US-08-774-104A-1

Perfect score: 1197

Sequence: 1 ATGATGACATATCAGCAAC.....ACTACATCATGTGATGA 1197

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_esthum21:\*  
64: em\_esthum22:\*  
65: em\_esthum23:\*  
66: em\_esthum24:\*  
67: em\_esthum25:\*  
68: em\_esthum26:\*  
69: em\_esthum27:\*  
70: em\_esthum28:\*  
71: em\_esthum29:\*  
72: em\_esthum30:\*  
73: em\_esthum31:\*  
74: em\_esthum32:\*  
75: em\_esthum33:\*  
76: em\_esthum34:\*  
77: em\_esthum35:\*  
78: em\_esthum36:\*  
79: em\_esthum37:\*  
80: em\_esthum38:\*  
81: em\_esthum39:\*  
82: em\_esthum40:\*  
83: em\_esthum41:\*  
84: em\_esthum42:\*  
85: em\_esthum43:\*  
86: em\_esthum44:\*  
87: em\_esthum45:\*  
88: em\_esthum46:\*  
89: em\_esthum47:\*  
90: em\_esthum48:\*  
91: em\_esthum49:\*  
92: em\_esthum50:\*  
93: em\_esthum51:\*  
94: em\_esthum52:\*  
95: em\_esthum53:\*  
96: em\_esthum54:\*  
97: em\_esthum55:\*  
98: em\_esthum56:\*  
99: em\_esthum57:\*  
100: em\_esthum58:\*  
101: em\_esthum59:\*  
102: em\_esthum60:\*  
103: em\_esthum61:\*  
104: em\_esthum62:\*  
105: em\_esthum63:\*  
106: em\_esthum64:\*  
107: em\_esthum65:\*  
108: em\_esthum66:\*  
109: em\_esthum67:\*  
110: em\_esthum68:\*  
111: em\_esthum69:\*  
112: em\_esthum70:\*  
113: em\_esthum71:\*  
114: em\_esthum72:\*  
115: em\_esthum73:\*  
116: em\_esthum74:\*

117: gb-gss18:\*  
118: gb-gss19:\*  
119: em-gss13:\*  
120: gb-gss20:\*  
121: gb-gss21:\*  
122: gb-gss22:\*  
123: gb-gss23:\*  
124: gb-gss24:\*  
125: em-gss14:\*  
126: em-gss15:\*  
127: em-gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	52.6	4.4	638	22	AM398687	AM398687 EST309187
2	47.6	4.0	351	23	AM596452	AM596452 s112e12.y
3	45.2	3.8	411	39	T42189	T42189 5452 Lambda
4	40.4	3.4	457	36	C71727	C71727 C71727 Rice
5	39.6	3.3	530	14	A1993966	A1993966 701494994
6	38.2	3.2	536	14	A1997116	A1997116 70152128
7	37.4	3.1	361	19	AV549266	AV549266 AV549266
8	37.4	3.1	558	19	AV560872	AV560872 AV560872
9	37	3.1	439	14	AL370502	AL370502 McBA58B12
10	36	3.0	280	30	BB345415	BB345415 BB345415
11	36	3.0	382	114	A2068859	A2068859 RPCI-23-4
12	35.8	3.0	261	32	BB535242	BB535242 BB535242
13	35.6	3.0	525	111	A0824936	A0824936 HS_5283_A
14	35.4	3.0	266	32	BB552509	BB552509 BB552509
15	35.4	3.0	286	30	BB373760	BB373760 BB373760
16	35.4	3.0	318	28	BB203673	BB203673 BB203673
17	35.4	3.0	445	40	W90867	W90867 m179e05.r1
18	35	2.9	925	121	CNS0091P	AL053013 Drosophila
19	33.8	2.8	442	2	AA272333	AA272333 vB2d02.r
20	33.8	2.8	712	88	AQ047433	AQ047433 CLM-1a3-u
21	33.4	2.8	433	23	AM582990	AM582990 1a08g11.y
22	33.4	2.8	463	36	D48678	D48678 R1C515048A
23	33.4	2.8	1101	121	CNS00EHN	AL069204 Drosophila
24	33.2	2.8	309	29	BB325541	BB325541 BB325541
25	33.2	2.8	360	12	A1648665	A1648665 tx63g12.x
26	33.2	2.8	509	112	AQ916714	AQ916714 nbeu067E
27	33.2	2.8	1101	121	CNS016NN	AL106573 Drosophila
28	33	2.8	450	7	A1008381	A1008381 EST202832
29	33	2.8	585	97	AQ418315	AQ418315 RPCI-11-1
30	33	2.8	916	123	CNS03YBE	AL266099 Tetradodon
31	32.8	2.7	300	36	C08171	C08171 C08171 yu11
32	32.8	2.7	360	16	AV189838	AV189838 AV189838
33	32.8	2.7	428	21	AW351529	AW351529 IL2-CT003
34	32.8	2.7	558	20	AW216289	AW216289 687047E01
35	32.8	2.7	668	25	AM850622	AM850622 IL3-CT021
36	32.6	2.7	266	32	BB540703	BB540703 BB540703
37	32.6	2.7	300	36	C36520	C36520 C36520 yu11
38	32.6	2.7	312	30	BB378305	BB378305 BB378305
39	32.6	2.7	322	38	BB233297	BB233297 BB233297
40	32.6	2.7	324	36	D71830	D71830 CELK082C8R
41	32.6	2.7	341	28	BB210509	BB210509 BB210509
42	32.6	2.7	344	28	BB210443	BB210443 BB210443
43	32.6	2.7	411	24	AM829046	AM829046 ra33d03.y
44	32.6	2.7	521	91	AQ305162	AQ305162 HS-2019_B
45	32.6	2.7	719	33	BE038450	BE038450 AA17F09.A

## ALIGNMENTS

RESULT 1

AM398687

LOCUS	AM398687	638 bp	mRNA	EST	07-FEB-2000
DEFINITION	EST309187 L. pennellii trichome, Cornell University Lycopersicon pennellii cDNA clone cLPT4016 5', mRNA sequence.				
ACCESSION	AM398687				
VERSION	AM398687.1 GI:6917157				
KEYWORDS	EST.				
SOURCE	Lycopersicon pennellii.				
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.				
REFERENCE	Alcala,J., Vredalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Mierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.				
AUTHORS	Generation of ESTs from wild tomato (Lycopersicon pennellii) trichomes				
TITLE	Unpublished (1999)				
JOURNAL	Contact: David Frisch				
COMMENT	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfisch@clemson.edu 5 prime sequence.				
FEATURES	Location/Qualifiers				
source	1..638 /organism="Lycopersicon pennellii" /db_xref="taxon:28526" /clone="cLPT4016" /clone_1ib="L. pennellii trichome, Cornell University" /issue_type="trichome" /dev_stage="mixed stages" /lab_host="SOL" /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Leaves of various stages were shaken in liquid nitrogen, shearing off trichomes. This procedure yielded a mixture of cells highly enriched for trichomes, with minor contamination by other types of leaf cells."				
BASE COUNT	179 a 126 c 164 g 169 t				
ORIGIN					
Query Match	4.4%; Score 52.6; DB 22; Length 638;				
Best Local Similarity	45.5%; Pred. No. 1.7e-05;				
Matches 187; Conservative	0; Mismatches 224; Indels 0; Gaps 0;				
OY	582	AGCAACACCGGAGCGTTAGCGCTATGTCAGGTCGTCAGTCGTCGCGGCGGTCGC	641		
DB	228	AACACCCCGAGCAGCTATTTTCATCGGACCGCCTCAGCATTTGTAGGCGTTTACG	287		
OY	642	TGGGACAGCGCGCGCGGAGCGATGCGTGCAGATATCGCGTTAGCGGAGCGTTGT	701		
DB	288	TGGAACAGCGGCTAGGCGTGCAGATGCGTGCAGATATTAATACATGATTTACT	347		
OY	702	AGCCAGTCACAAAACCGCATTTGACATCATATGTGTGGCAGATTCACGACGCTT	761		
DB	348	ATCCAGGCTTAAGGACCATATGATTTGCTATAGTCGGAGTGCATTAAGAGAAATTT	407		
OY	762	GCAAGACGTACAGCAGCGCTTAAGGTCGCGGCGCATGCCAAGTTTACTCAAAATAGCA	821		
DB	408	GGAGGCGTGTGTTCAAGCGTTTATTTGAACCAAGAAACCAATTAAGAAATTTTACAG	467		
OY	822	AGTTCAACATTTTACACACCAATTACAGGAGGACATTCGCGACATTTAAGTGTACCG	881		
DB	468	AGTTCAACATTTTACAGCGTGTGATTTAGAGGGGAGACCTCAGGCTGAAGATGATGATT	527		
OY	882	GATTGTTGACCGCTTGATTCACACGACGACGATGGGCGTCCACGATGACGCGCTT	941		
DB	528	GATTATGTCACTGCGTGCACCGCTACCTCAGCAGTTGTGGTATCTTACAGAAATGACG	587		

[illegible][illegible]



Dd	497	TGCATTCGAACCTCCACTGTTTGTGGCGCTTCCAGCAAGAAGCAAGCGTTTGATTAAAG	438
Oy	936	CCCATGAGAACACCCTCGTGCTTTGTTCGAGAGTCTATTGGCTATTTTAACCGCAGATA	1015
Dd	437	AGATATCATTCATTCATGAGGAATGTATGGGACCTATTGATTTTTTGTGGCGAGG	378
Oy	1016	ATATGTC 1021	
Dd	377	AGATGTC 372	
RESULT 6	A1997116/c		
LOCUS	A1997116	536 bp	mRNA EST 08-SEP-1999
DEFINITION	701552128 A. thaliana, Columbia Col-0, root-2 Arabidopsis thaliana		
ACCESSION	CDNA clone 701552128, mRNA sequence.		
VERSION	A1997116		
KEYWORDS	A1997116.1 GI:5844021		
SOURCE	EST.		
ORGANISM	thale cress.		
REFERENCE	Arabidopsis thaliana		
AUTHORS	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta: eudicotyledons: core eudicots: Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi-		
	s (bases 1 to 536)		
	Chen, Y., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Monahanlouca, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nodriga, A., Murry, L., Turner, C., Kirkorian, S., Elder, L. and Hanson, D.		
TITLE	Arabidopsis thaliana Gene Expression Microarray		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 463 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.		
FEATURES	location/qualifiers		
source	1..536		
	/organism="Arabidopsis thaliana"		
	/cultivar="Columbia Col-0"		
	/db_xref="taxon:3702"		
	/clone="701552128"		
	/clone_lib="A. thaliana, Columbia Col-0, root-2"		
	/tissue_type="root"		
	/dev_stage="4 - 7 weeks"		
	/note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA library was derived from untreated root tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was bluntended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."		
BASE COUNT	159 a	119 c	91 g 145 t 22 others
ORIGIN			
Query Match	3.2%; Score 38.2;	DB 14;	Length 536;
Best Local Similarity	45.4%;	Pred. NO. 0.57;	Mismatches 113;
Matches	94;	Conservative 0;	Gaps 0;
Oy	62	CACCTGAAACAACACTGAATTTGGCGCTTACTTATTCGACACCCGCTGATGATACCTTAGCTT	121
Dd	393	CATTCACANNNNNNNNNNNNNNNNAATGTTCCGACACTAAGTATGATGCTTTGACATC	334
Oy	122	TTCGCATTGGCCCAATCGCTACGCGAAAAAGCGCTCAGGCATTACAGAGTGGCGTTGTTT	181

Db	333	GTCGTGCCATGCTCACTTCTCCTCCAAAAACCTGCCAATTAAAGACATGACTGTGTGCA	274
Oy	182	TTTGCGCGAGTCATTGGTAGTAACAAGAGTACCAGCCAGTCGTGAATTGGATGGCGGTTTTT	241
Db	273	TTTGTGTGGATGCACCATCTGAAGCAGTGTGTGTGCCGTGTGACATGTGCCGGATGCA	214
Oy	242	GGTTGTGTCCCCGAAGTATGTGTACCA	268
Db	213	TTTCTTGTGTTGAAGAAGATTGAAAACA	187
RESULT	7		
LOCUS	AV549266	361 bp	mRNA EST 23-JUL-2000
DEFINITION	AV549266 Arabidopsis thaliana roots Columbia Arabidopsis thaliana		
ACCESSION	CNDA clone R205e08r 5', mRNA sequence.		
VERSION	AV549266		
KEYWORDS	AV549266.1 GI:8720679		
SOURCE	EST.		
ORGANISM	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 361) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. A large scale analysis of cdna in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cdna libraries DNA Res. 7, 175-180 (2000) Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizue@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/ Location/Qualifiers 1..361 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="R205e08r" /clone_1lb="Arabidopsis thaliana roots Columbia" /tissue_type="Roots" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"		
JOURNAL	DNA Res. 7, 175-180 (2000)		
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizue@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/ Location/Qualifiers 1..361 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="R205e08r" /clone_1lb="Arabidopsis thaliana roots Columbia" /tissue_type="Roots" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"		
FEATURES	source		
BASE COUNT	103 a 72 c 75 g 111 t		
ORIGIN			
Query Match	3.1%; Score 37.4; DB 19; Length 361;		
Best Local Similarity	48.8%; Pred. No. 0.89;		
Matches 101; Conservative	0; Mismatches 100; Indels 0; Gaps 0;		
Oy	62 CACTTGAACAACGTAATTTGGCGCTTACTTTGGACACCCCGCTGATGATTAATTAACGTT	121	
Db	22 CATCTCAAGCACCAACACACACCAACCAATGTTCTGCACTAAGTAGTATGATCTTGATCTC	81	
Oy	122 TTGGCATTTGGCCCATTCGTCAGCGCAAAAACGGCTCAGGCAATTACAAGTGGCGTTGTT	181	
Db	82 GTGTGCGCATGTGACTCTCTCTCCCAAACGCGCAATTAAGAGATGAGCTGTGTGTGA	141	
Oy	182 TTGGTGGCAGTCATTTGTTGAACAAGATTAACCCGAGTCGTGAATTTGATGGCGGTTTTT	241	
Db	142 TTTGTGTGATGCACCATCTGAAGCAGTGTGTGTGCGCTGTGGACATGTGCCCGATGCA	201	
Oy	242 GGTTGTGTCCCAGAGTATGTGTACCA	268	
Db	202 TTTCTTGTGTTGAAGAAGATTGAAAACA	228	
RESULT	8		
LOCUS	AV560872/c	558 bp	mRNA EST 23-JUL-2000

[illegible]

BASE COUNT	ORIGIN	FEATURES	SOURCE	TITLE	COMMENT				
137	a	74	c	130	g	98	t	Medicago truncatula ESTs from nitrogen-starved roots	Unpublished (2000)
								Genoscope - Centre National de Sequencage	Contact: Genoscope
								BP 191 91006 Evry cedex - France	Email: segr@genoscope.cns.fr, Web : www.genoscope.cns.fr
								Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de	Biologie Moleculaire des Relations Plantes-Microorganismes,
								CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :	M-et@toulouse.inra.fr Website :
								http://sequence.toulouse.inra.fr/Mtruncatula.html).	Location/Qualifiers
								1. 439	
								/organism="Medicago truncatula"	
								/cultivar="Jemalong"	
								/db_xref="taxon:3880"	
								/clone="MTBA38B12"	
								/clone_lib="MtBA"	
								/tissue_type="root tips"	
								/dex_stage="harvested after 3 days of N-starvation"	
								/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2:	
								XhoI; Plants were grown in an aeroponic chamber for 14	
								days on nitrogen-rich medium followed by 3 days on N-free	
								medium. RNA was extracted from root tips (1-3 cm). cDNA	
								was prepared from polyA+ enriched RNA. The cDNA was	
								directionally ligated into Uni-zapR vector from	
								Stratagene and packaged using gigaPack Gold packaging	
								extracts. Plasmids containing cDNA inserts were	
								mass-excised from phage stocks using Exasit helper phage	
								and propagated in SOUR cells. Clone ordering and	
								sequencing was performed by the Centre National de	
								Sequencage (Genoscope, Evry, France)."	

[illegible]



Y., Shigenoto, Y., Shinaagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiunga, N., Toyota, T., Tsunoda, Y., Matahiki, A., Matanabe, S., Yamamura, T., Yamataka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse (2005)  
Unpublished (2005)  
Contact: Yoshihide Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-reserc.riken.go.jp,  
URL: <http://genome.rtc.riken.go.jp/>  
Carinci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasakita, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)  
Itoh, M., Kitsunishi, T., Akiyama, K., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carinci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. *Genome Res.* 9 (5), 463-470 (1999)  
Carinci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details

Query Match	3.0%;	Score 36;	DB 30;	Length 280;
Best Local Similarity	62.0%;	Pred. No. 2.3;		
Matches	57;	Conservative	0;	Mismatches 35;
			Indels	0;
			Gaps	0;

07 ATTACATTGGCAGCCCATGAGAAGACACCTCGTGGCTTTTGGCAGGCTCTATTGGCTATT 1003  
11 ||||| - - ||||| ||||| ||||| - - |||||  
Db 182 ATACATTGGACACCAACTGCCAAGACACTCCTGGCGTTACTTTAGTGTGCTTAGTATT 241

[illegible]

458 TTGGCGGCACAGACCGTCGAGTTATCCGACAGCTTGACACTGGCACAAATTTATT 513  
 70 CTATCAATTCACAGTACACACATTTTAAAGAAAGCGCAGACTGAAAAAATAATT 15





TITLE  
JOURNAL  
COMMENT

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshitake Hayashizaki

Genome Exploration Research Group, Life Science Tsukuba Center,

The Institute of Physical and Chemical Research (RIKEN), Genomic

Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: +81-298-36-9013

Fax: +81-298-36-9098

Email: genome-res@rtc.riken.go.jp,

URL: http://genome.rtc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,

N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Thermostabilization and thermocatalysis of thermostable enzymes by

trehalose and its application for the synthesis of full length

CDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,

Y., and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P., and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

FEATURES  
Source

Location/Qualifiers

1..286

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="C130071N21"

/clone\_lib="RIKEN full-length enriched, 16 days embryo

head"

/sex="mixed"

/tissue\_type="head"

/dev\_stage="16 days embryo"

/lab\_host="DH10B"

/note="Site 1: Sali; Site 2: BamHI: cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGATCCCAAGACCTCTTTTCTTTTCTTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGATCCCAAGATTAATTAATCCCCCCCCCC 3']. cDNA

was cloned into the XhoI and BamHI sites. Vector: a

modified pBluescript KS(+) after bulk excision from Lambda

FLC I"

BASE COUNT 90 a 40 c 39 g 117 t  
ORIGIN

Query Match

Best Local Similarity 3.0%; Score 35.4; DB 30; Length 286;

Matches 66; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 1078 ACTTATTTGCGGCGGCGGATGTGCTGACTCCGATCGGCAACAGATATGAAGAA 1137

DB 273 AATTATTTTGTGCTGATATCTTTGTTAAATGAAGCTCTAGCTGAATTAAGGGTTT 214

QY 1138 ACTGGGTGAATTTGACCAATGCGCAATGTTTAAGAGTACATCATGTGAA 1194

DB 213 TTAGTGTGTAATTTTACCAATGACCAATTAATTAATTTTACATCTTTGGAA 157



**THIS PAGE BLANK (user)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 01:23:18 ; Search time 55.47 Seconds  
(without alignments)  
244.726 Million cell updates/sec

Title: US-08-774-104A-2  
Perfect score: 2009  
Sequence: 1 MMTYHETRALAQSDLQQLYA.....ETGLKFEPRQLLDYNNIVE 397

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues  
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
A.Geneseq.36:.\*  
1: /SID6/gcgdata/geneseq/AA1980.DAT.\*  
2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2009	100.0	397	19	W69162	Human neutral sph
2	360	17.9	580	20	Y50340	C. roseus ICS prot
3	344	17.1	591	20	Y50335	E. coli entC prote
4	329.5	16.4	271	20	V33698	E. coli entC prote
5	302.5	15.1	391	20	Y50336	P. fluorescens ICS
6	297	14.8	331	20	Y50338	P. fluorescens ICS
7	209	10.4	491	20	V41302	Tobacco anthranila
8	195.5	9.7	603	20	V42112	Corn anthranilate
9	195	9.7	616	20	V41301	Tobacco anthranila
10	189.5	9.4	606	20	W93814	Rice anthranilate
11	186.5	9.3	446	20	V42113	Corn anthranilate
12	184.5	9.2	615	18	W26685	Maize C28 allele a

13	183.5	9.1	615	18	W26686	Maize anthranilate
14	181.5	9.0	577	20	W93815	Rice ASA first iso
15	181.5	9.0	577	20	W93810	Rice anthranilate
16	170	8.5	518	8	P70506	Type enzyme. P70
17	153	7.6	401	18	W55536	H. pylori ORF 07ap
18	153	7.6	567	18	W55703	H. pylori ORF 07ec
19	152	7.6	420	18	W55418	H. pylori ORF hp2e
20	150	7.5	193	9	P80877	Hook region #8 con
21	150	7.5	262	5	P40023	Fusion protein #8 con
22	150	7.5	405	4	P30202	Sequence encoded b
23	149	7.4	559	19	W98537	H. pylori GPO 148
24	145	7.2	225	8	P70359	Sequence encoded b
25	145	7.2	225	8	P70061	Sequence encoded b
26	142	7.1	406	4	P30203	Sequence of human
27	140	7.0	609	4	P30206	Sequence encoded b
28	137	6.8	529	20	Y29293	Deduced amino acid
29	107.5	5.4	1224	18	W26735	Staphylococcus car
30	97	4.8	2532	20	Y33729	Photobacterium lum
31	96.5	4.8	1346	21	Y77195	S. venezuelae macr
32	96.5	4.8	1346	21	Y67204	Nardoniella syntha
33	96.5	4.8	1219	21	Y77180	S. venezuelae pik
34	96	4.8	1224	20	Y36851	Protein involved i
35	95.5	4.8	600	20	W82660	Cauliflower L-gala
36	94.5	4.7	1374	19	W72225	HSV-2 strain SB5 C
37	94.5	4.7	1384	19	W72224	HSV-2 strain SB5 C
38	94.5	4.7	1386	19	W72117	HSV-2 strain SB5 C
39	94.5	4.7	1366	19	W72039	HSV-2 strain SB5 C
40	94	4.7	1092	19	W41602	Staphylococcus epl
41	94	4.7	3739	21	Y77193	S. venezuelae macr
42	94	4.7	3739	21	Y77201	S. venezuelae pik
43	94	4.7	3739	21	Y67202	Nardoniella syntha
44	93.5	4.7	991	21	Y83171	Cell wall protein
45	93.5	4.7	991	21	Y70120	Staph. epidermidis

## ALIGNMENTS

RESULT	1	
W69162	W69162 standard; Protein; 397 AA.	
ID	W69162;	
AC	20-OCT-1998 (first entry)	
DT	Human neutral sphingomyelinase.	
XX	Human: neutral sphingomyelinase; N-Smase; diagnosis; Crohn's disease;	
XX	obesity; diabetes; Alzheimer's disease.	
KW	Homo sapiens.	
OS		
XX	Key	Location/Qualifiers
FH	Misc-difference 59	/note="encoded by GTTCTT"
FT		
XX		
PN	W09828445-A1.	
PD	02-JUL-1998.	
XX		
PF	23-DEC-1997;	97MO-US24051.
XX		
PR	24-DEC-1996;	96US-0774104.
XX		
PA	(UYJO ) UNIV JOHNS HOPKINS.	
XX		
PI	Chatterjee S;	
XX		
DR	WPI; 1998-377673/32.	
XX	N-PSDB; V41196.	
XX		
PT	Human neutral sphingomyelinase - used to, e.g. treat N-Smase	

PT associated disorders, e.g. Crohn's disease, obesity, diabetes, and  
PT Alzheimer's disease  
XX  
XX  
XX Claim 19; Page 29-30; 47pp; English.  
XX  
CC The present sequence represents human neutral sphingomyelinase (N-SMase).  
CC A host cell containing a vector comprising a nucleotide sequence  
CC encoding N-SMase can be used to produce N-SMase. N-SMase can be used in  
CC a method for identifying a compound useful in the diagnosis or treatment  
CC of a human neutral sphingomyelinase related disorder. N-SMase, and a  
CC nucleotide sequence encoding N-SMase, can be used for modulating N-SMase  
CC activity, and for treating a disorder associated with N-SMase. The  
CC N-SMase disorders that can be treated with the nucleotide sequence or  
CC N-SMase, is an inflammatory disorder, arthritis, osteoarthritis, Crohn's  
CC disease, obesity, diabetes, cirrhosis, susceptible tumours, central  
CC nervous system disorder, vascular stenosis, arterial occlusion arising  
CC from plaque formation, cardiac disease where LV dysfunction occurs,  
CC hypercholesterolaemia, cholesterol ester storage disorder, renal failure,  
CC HIV infection, depression, schizophrenia, neurodegeneration, and  
CC Alzheimer's disease. An antibody against N-SMase can be used to reduce  
CC tumour necrosis factor alpha (TNF-alpha) induced apoptosis of mammalian  
CC cells.  
XX  
XX Sequence 397 AA;  
SQ

Query Match 100.0%; Score 2009; DB 19; Length 397;  
Best Local Similarity 100.0%; Pred. No. 8,6e-181;  
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MMTYHETRALQOSDLOQLYALLETTEFGAYRTPADDTLRFGIGAIATAKTAQALQGAVF 60  
DB 1 mmyhetralqgsdlqglyaaletefgyafatpaddtlrfjgatalaktaqalgavf 60  
OY 61 GAOSFDEQEXYQSELMAGFWFVPEVWVTIADKIRFGSDTVSDFTTWIAQFYPKQPNY 120  
DB 61 gagsfdeeqyqselmagfwfvpewvvtiaadkirtfsgdsvdfttcwlaqtpkqpnvt 120  
OY 121 TSHVTDVDEWIERENLIDTLAIDQTLAKVVGROQTQLSDTLRLAQIIRALAEQANTY 180  
DB 121 tshvtdvdewtierenlidtlaidqtlakvvgfgrqtlqsdltlrqlaqliralaeanqy 180  
OY 181 HVTLKRHDELISATPERVLVMSGQIATPAVAGTSRRGTGADGADIALGAILASOKNRI 240  
DB 181 hvtlkrhdelisatpervlvmsgqiatpavagtsrrgtgdadialgeailasqknri 240  
OY 241 EHOVVASITRIRDVTSKLVPAWPSLLKNQVOHLTPITGDIAAHLSTVATYDRALHP 300  
DB 241 ehovvasitrlrdvtsklvpawpsllknqvohlytpitgdiaahlstvaydrhlhp 300  
OY 301 TPALGVPREALYYIATHEKTPRGLFAGPIGFTADNSGEFVVGIRSMYVNOTORRATL 360  
DB 301 tpalgvpreealyyiathektrpgrlfgapigftadnsgefvvgirsmynvnotqrratl 360  
OY 361 FAGAGIVADSDAQOEYETGLKFEPMROLKDYNHVE 397  
DB 361 fagagivadsdaqeeyetglkfepmqllkdynhve 397

RESULT 2  
Y50340  
ID Y50340 standard; Protein; 580 AA.  
XX  
XX Y50340;  
XX  
DT 14-JAN-2000 (first entry)  
DE  
XX C. roseus ICS protein.  
XX  
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KM pathogen inducible promoter; antipathogenic protein; toxin;  
KM antifungal protein; albumin-type protein; hypersensitive response.  
XX

OS Catharanthus roseus.  
XX  
XX WO950423-A2.  
XX  
XX 07-OCT-1999.  
XX  
XX 25-MAR-1999; 99WO-EP02176.  
XX  
XX 31-MAR-1998; 98US-0080203.  
XX  
XX 03-APR-1998; 98US-0080625.  
XX  
XX (MOGE-) MOGEN INT NV.  
XX  
XX (UYLE-) RIJSCUNIV LEIDEN.  
XX  
XX (UYNI-) UNIV NIJMEGEN.  
XX  
XX Linhorst HJM, Verpoorte R, Verberne MC, Moreno PRH;  
PI Van Tegelen LJP, Willems GJ, Groes AF, Stuijver MH, Custers J;  
PI Simons LH, Melchers LS, Bol JF;  
DR WPI; 1999-610856/52.  
XX  
XX N-PSDB; 223758.  
XX  
XX Method for inducing pathogen resistance in plants  
PT  
XX  
XX Claim 11; Page 63-64; 66pp; English.  
XX  
XX  
XX This invention describes a novel method for the induction of pathogen  
XX resistance in plants, by transformation with an expression cassette  
XX harboring a gene coding for an isochlorismate synthase (ICS). A pathogen  
XX inducible promoter can be used to drive expression of a heterologous  
XX protein. The heterologous protein used in the method of the invention is  
XX an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
XX saccharinase, or antifungal proteins isolated from *Mirabilis jalapa*,  
XX *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,  
XX *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Citrus*,  
XX *Lactyrus*, *Cilicoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
XX proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
XX gliadin and wheat-alpha-amylase, or a protein that can induce a  
XX hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and  
XX N-protein from tobacco. This sequence represents the *Catharanthus roseus*  
XX isochlorismate synthase ICS protein which is described in the method of  
XX the invention.  
SQ

Sequence 580 AA;  
SQ

Query Match 17.9%; Score 360; DB 20; Length 580;  
Best Local Similarity 29.6%; Pred. No. 1,6e-25;  
Matches 107; Conservative 65; Mismatches 157; Indels 32; Gaps 10;

OY 60 FGAQSFDEQEXY--PQSELMAGFWF-VPEV-----WVTIAD-----K 93  
DB 216 ygaltrfdarphlapewkafgsfymypqvefdelngssmlaatvwdnalsltqgalvr 275  
OY 94 ITFGSDTVSDFTTWIAQFYPKQPNYVTSVYDEVWIERENLIDTLAIDQT-LAKVVF 152  
DB 276 lqtmegysvstsvksrlrgdv-shstlsvskanlpdrctswdltlrvleelgnksypltkvvl 334  
OY 153 GRQ-OTLOISDTLRLAQIIRALAEQANTYHVVYLRKHD-ELFISATPERVLVMSGQIATA 210  
DB 335 arrgsqvltsdldiplawssfkadgkdayqfclqheapafignpeqlfgrdqllvtsie 394  
OY 211 AVAGTSRRGTGADGADGADIALGAILASOKNRIEHOVVASITRIRDVTSKLVPAWPSLLK 270  
DB 395 alaaatrargesdldlqnahdlfsspkdnheiaivrenrtqldalcsvecepmksvyrk 454  
OY 271 NKQVOHLTPITGDIAAHLSTVATYDRALHPALGVPREALYYIATHEKTPRGLFAGP 330  
DB 455 lkrighlyarfaragtrrseddefklsslhprpavvgffmedarkfiaeemfdrllyagp 514  
OY 331 IGFPTADNSGEFVVGIRSMYVNOTORRATLRFAGAGIVADSDAQOEYETGLKFEPMROL 390  
DB 515 vgfifgags-dfsvyglrsallgk-dagallyyglvvegsdpalewgelelkaagfmklm 572



```
OY 391 K 391
Db 573 K 573

RESULT 3
Y50335
ID Y50335 standard; Protein; 391 AA.
AC Y50335;
DE 14-JAN-2000 (first entry)
XX
XX E. coli entc protein.
XX
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin; entc;
KM antifungal protein; albumin-type protein; hypersensitive response.
XX
OS Escherichia coli.
XX
XX WO9950423-A2.
XX
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
PR 03-APR-1998; 98US-0080625.
XX
XX (MOGE-) MOGEN INT NV.
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (UYNI-) UNIV NIJMEGEN.
XX
XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
PI Simons LH, Melchers LS, Bol JF;
DR WPI; 1999-610856/52.
DR N-PSDB; 223756.
XX
XX Method for inducing pathogen resistance in plants -
PT
XX
XX Example 3; Page 51-53; 66pp; English.
XX
XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magainins, toxins from Bacillus
CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
CC Amaryanthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Cnicus,
CC Lathyrus, Clitoria, Allium seeds, Avena and Impatiens and albumin-type
CC proteins, such as thionin, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat-alpha-amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and
CC N-protein from tobacco. This sequence represents the Escherichia coli
CC isochorismate synthase entc gene which is described in the method of
CC the invention.
XX
XX Sequence 391 AA;
SQ
```

```
Query Match 17.1%; Score 344; DB 20; Length 391;
Best Local Similarity 26.6%; Pred. NO. 2.9e-24;
Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;
```

```
OY 19 YAALETTEFGAVFATPA-----DDTLRFGICGATTAATAQAAGAVF-GAQSFEQDEYP 71
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 27 yrstfscgarfddepavngsdpspfqkilaalfadakagqikpvmvgaipfprq p 85
```

```
OY 72 OSELMAGFVPEVWMTTIAADKITFGSDTVSDFTTTLAOFVPRKQNTVTSHTDEVDM 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 86 ss-----Lyp-----sw-gsfargkqasarrftr----- 111

OY 132 ERTENLIDPLAT-DQTLAKVFGROQTQLDS--DTLRALQILRALAEQANTYHYVLRK- 186
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 112 sgslnvvergaipqetlffegmvaraaltatcpvdkvslrilditldaidsyvlert 171

OY 187 -----HDEL-----FISATPERLVAMSGGQATATAVAGTSRGTGDADIALGEA 231
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 172 laqnpvsynfhvpladgavlllgaspe11lrkdgerfssiplagsarrqpdvldreagr 231

OY 233 LLASQKNRIEHOYVVAATITRLQDYVTSIAKVAMPSLKNKOYOHLYPTITGDIAHLSV 291
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 232 llaekdrhelvpcamkevltresse1hvpsspq1ltlplwhatpfdgkxansqena 291

OY 292 TAIIVRLHPTPLAGVPRRAALXYIATHEKTPRGAFIGYFTADNSGEFVGTIRSMV 351
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 292 ltlac1hptpalsgfpnqaatqvlaelepfare1fg91vgwcdeegngewvltlrcak 351

OY 352 NOTORRATLFAAGAGIVADSDAQOEXEETGLKPEPM 386
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 352 renqvr--l1agag1vpaspslgewretgkx1stm 384
```

```
RESULT 4
Y33698
ID Y33698 standard; Protein; 271 AA.
AC Y33698;
DE 14-JAN-2000 (first entry)
XX
XX E. coli entc protein fragment.
XX
XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin; entc;
KM antifungal protein; albumin-type protein; hypersensitive response,
XX Escherichia coli.
XX
XX WO9950423-A2.
XX
XX 07-OCT-1999.
XX
XX 25-MAR-1999; 99WO-EP02176.
XX
XX 31-MAR-1998; 98US-0080203.
PR 03-APR-1998; 98US-0080625.
XX
XX (MOGE-) MOGEN INT NV.
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (UYNI-) UNIV NIJMEGEN.
XX
XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
PI Simons LH, Melchers LS, Bol JF;
DR WPI; 1999-610856/52.
DR N-PSDB; 223756.
XX
XX Method for inducing pathogen resistance in plants -
PT
XX
XX Example 3; Page 53-54; 66pp; English.
XX
XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magainins, toxins from Bacillus
```

CC thuringiensis, or antifungal proteins isolated from *Mirabilis jalapa*,  
CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*,  
CC *Lathyrus*, *Citrovia*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
CC hypersensitive response, such as Cf, B53 and Po proteins from tomato and  
CC N-protein from tobacco. This sequence encodes the *Escherichia coli*  
CC isochorismate synthase *entC* gene which is described in the method of  
CC the invention.  
CC  
XX

Sequence 271 AA;

Query Match 16.4%; Score 329.5; DB 20; Length 271;  
Best Local Similarity 32.5%; Pred. No. 3.9e-23;  
Matches 91; Conservative 46; Mismatches 124; Indels 19; Gaps 5;

QY 110 QFVFKQDNVTYTHSVNDEVMIEKTEMLIDTLAIDQLKAVFGRQDTQLSDTLRLAQI 169  
DB 1 qalpeq-----lfeqmvareaalatpgvd---kvvlslridltddaidsqvl 47

QY 170 IRLALAQ---ANRYHVVLKRRHDELISATPERLVAMSGGQIATAAAGTSRGTGADDI 226  
DB 48 lertlaqnpvsynfthvpl-adgavllqaspepllirkdgeritssiplagsarrqpdevldr 106

QY 227 ALGELLASQKNRIHQYVVASITRLODYTTSLKVPAMPSLKNKOVOHLPTITGDIA 286  
DB 107 eagrllasekdrhehelveqamkevlrersselhyvpsqdlitpplwhatpfegkan 166

QY 287 AHSVTAIVRLHPTPLAGVPREALYITATHEKTRGIFACIGFTYDNGSEFYVGI 346  
DB 167 sgenatlacilhpplpsgfpbqaatqvlaeplefdrrelfvgivgcdeegngevvtli 226

QY 347 RSMVNOTORATLTFAGAGIVADSDAQOEYENGKKEPM 386  
DB 227 rcaiklengvtr--lfaagavlpaspslgwregeykvistm 264

RESULT 5  
Y50336 standard; Protein; 391 AA.

AC Y50336;  
DT 14-JAN-2000 (first entry)

DE P. fluorescens ICS orfa protein.

XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KW pathogen inducible promoter; antipathogenic protein; toxin;  
XX antifungal protein; albumin-type protein; hypersensitive response.

OS Pseudomonas fluorescens.

XX WO950423-A2.

XX 07-OCT-1999.

XX 25-MAR-1999; 99WO-EP02176.

XX 31-MAR-1998; 98US-0080203.

XX 03-APR-1998; 98US-0080625.

XX (MOGE-) MOGEN INT NV.

XX (UYLE-) RIKSUNIV LEIDEN.

XX (UYNI-) UNIV NIJMEGEN.

XX Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;

XX Van Tegelen LJP, Mullens GJ, Groes AF, Stulver MH, Custers J;

XX Simons LH, Melchers LS, Bol JF;

XX WPI; 1999-610856/52.

XX N-PSDB; Z23757.

XX Method for inducing pathogen resistance in plants  
PT Example 3; Page 54-56; 66pp; English.

XX This invention describes a novel method for the induction of pathogen  
CC resistance in plants, by transformation with an expression cassette  
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen  
CC inducible promoter can be used to drive expression of a heterologous  
CC protein. The heterologous protein used in the method of the invention is  
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,  
CC saccharide oxidase, oxalate oxidase, magalins, toxins from *Bacillus*  
CC *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,  
CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Cnicus*,  
CC *Lathyrus*, *Citrovia*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type  
CC proteins, such as thionine, napin, barley trypsin inhibitor, cereal  
CC gliadin and wheat-alpha-amylase, or a protein that can induce a  
CC hypersensitive response, such as Cf, B53 and Po proteins from tomato and  
CC N-protein from tobacco. This sequence represents the *Pseudomonas*  
CC fluorescens isochorismate synthase *orfa* protein which is described in  
CC the method of the invention.  
CC  
XX

Sequence 391 AA;

Query Match 15.1%; Score 302.5; DB 20; Length 391;  
Best Local Similarity 28.7%; Pred. No. 2.3e-20;  
Matches 107; Conservative 46; Mismatches 157; Indels 63; Gaps 10;

QY 68 QEYPOSELMAGFWFV---EYVVTIADKITYF---GSDTVSDFTTWLAQ----- 110  
DB 12 eevqlaevqrsfstsgdrelavqgmrlgetpaldgdanslfqqlaqaldrareeqg 71

QY 111 -----FVFKQPNVTYTHSVNDEVMIEKTEMLIDTLAID 144  
DB 72 snpllvgaipdpaeapcl---yipenaqvtrtdihaktgmslpelegknhdpqafk 128

QY 145 QTLAKVFG-----ROOTLQSDTLRLAQIIRALAEQANT-----YHVVLKRRH-EL 190  
DB 129 ravehavvnfnhsdvtrkavlsyqrelifandvdtsalqhnkqnpqsyhtrvmpdgctt 188

QY 191 FISATPERLVAMSGGQIATAAAGTSRGTGADADIALGELLASQKNRIHQYVVASIT 250  
DB 189 ligvspelivrkeglsslnplagsakrmadpeadrnadallsekdnyhgtvqdlv 248

QY 251 TRLODYTTSLKVPAMPSLKNKOVOHLPTITGDIA-AHSVTAIVRLHPTPLAGVPR 309  
DB 249 sqglkictqlnvppqrpalsltpalwhstirlegtladpavsalqlacrlnpbcvqgfp 308

QY 310 EAALYVITATHEKTRGIFACIGFTYDNGSEFYVIRSMVNOTORATLTFAGAGIVAD 369  
DB 309 eratrrllrfvepferygltgmvgwcdagngewvvtlircglvtrnkvr--lfaagavle 366

QY 370 SDAQOEYETGLK 382  
DB 367 sspdsawaevqtk 379

RESULT 6  
Y50338 standard; Protein; 331 AA.

AC Y50338;

DT 14-JAN-2000 (first entry)

DE P. fluorescens ICS orfa protein fragment.

XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;  
KW pathogen inducible promoter; antipathogenic protein; toxin;  
XX antifungal protein; albumin-type protein; hypersensitive response.

OS Pseudomonas fluorescens.

```
XX PN MO950423-A2.
XX PD 07-OCT-1999.
XX PF 25-MAR-1999; 99MO-EP02176.
XX PR 31-MAR-1998; 98US-0080203.
XX PR 03-APR-1998; 98US-0080625.
XX PA (MOGE-) MOGEN INT NV.
XX PA (UYLE-) RIJKSUNIV LEIDEN.
XX PA (UYNI-) UNIV NIJMEGEN.
XX PI Linthorst HJM, Verpoorte R, Verberne MC, Moreno PRH;
XX PI Van Tegelen LJP, Willems GJ, Croes AF, Stuiver MH, Custers J;
XX PI Simons LH, Melchers LS, Bol JF;
XX DR WPI: 1999-610856/52.
XX DR N-PSDB: Z23757.
XX PT Method for inducing pathogen resistance in plants
XX PS Example 3: Page 58-59; 66pp; English.
XX CC This invention describes a novel method for the induction of pathogen
XX CC resistance in plants, by transformation with an expression cassette
XX CC harboring a gene coding for an isochlorismate synthase (ICS). A pathogen
XX CC inducible promoter can be used to drive expression of a heterologous
XX CC protein. The heterologous protein used in the method of the invention is
XX CC an anti-pathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
XX CC saccharite oxidase, oxalate oxidase, magainins, coxins from Bacillus
XX CC thuringiensis, or antifungal proteins isolated from Mirabilis jalapa,
XX CC Amaranthus, Raphanus, Brassica, Sinapis, Arabidopsis, Dahlia, Nicotiana,
XX CC Lathyrus, Cicer, Allium seeds, Avena and Impatiens, and albumin-type
XX CC proteins, such as chitinase, napin, barley trypsin inhibitor, cereal
XX CC gliadin and wheat-alpha-amylase, or a protein that can induce a
XX CC hypersensitive response, such as Cf. Bs3 and Po proteins from tomato and
XX CC N-protein from tobacco. This sequence represents a Pseudomonas
XX CC fluorescens isochlorismate synthase orfa protein fragment which is
XX CC described in the method of the invention.
XX SO Sequence 331 AA;

Query Match 14.8%; Score 297; DB 20; Length 331;
Best Local Similarity 32.2%; Pred. No. 6e-20;
Matches 87; Conservative 43; Mismatches 134; Indels 6; Gaps 5;

QY 116 PNVTTSHTVDEVMWIERENLIDTLAIDOTLAKVFGROOTLSDTLRLAQTIRAL-A 174
DB 53 pellegknipdegafravehavvfr-hsdvkravlsvqrellifandvdsalqhnika 111
QY 175 EOAMTVHVLKRHD-ELFISATPERLVAMSGQIATAVAGSRGTCGADIDALGELL 233
DB 112 gnpesgylfrmpdgctclivspellivkeglsinsplagsakmadpeadrinaadwli 171
QY 234 ASQNRLEHOYVASTITRLQDVTTLKVPAMPSLKKNQVOHLYTPTTGDA-AHLSTV 292
DB 172 tsekdhghegftvqtdivsglqcltqlnvqprslstspalwhlstrlegtladpavsaal 231
QY 293 AIVDRLHPTPLGVPREALYYIATHEKTPRGAGPIGFTTADNSEEFGVIGSMYVN 352
DB 232 qlactilnpbcavcgfpfctarrllirfvepfertgltgmvgcdagngewvllircgvtvr 291
QY 353 OTORRATLFAAGIYVADSDAQOEVEETGLX 382
DB 292 tnkvrlfagaglyveasspdseawevgtk 319

RESULT 7
Y41302
ID Y41302 standard; Protein; 491 AA.
```

```
XX AC Y41302;
XX DT 18-JAN-2000 (first entry)
XX DE Tobacco anthranilate synthase ASAI.
XX KW Promoter; tobacco; anthranilate synthase; tolerance; analogue; Trp;
XX KW tryptophan; transgenic plant.
XX OS Nicotiana tabacum.
XX PN US5965727-A.
XX PD 12-OCT-1999.
XX PF 31-DEC-1997; 97US-0001826.
XX PR 26-JUL-1996; 96US-0025140.
XX PR 25-JUL-1997; 97US-0937739.
XX PA (UNII ) UNIV ILLINOIS FOUND.
XX PI Withholm JM, Brotherton JE, Song H;
XX DR WPI: 1999-579943/49.
XX DR N-PSDB: Z30446.
XX PT Plant tissue culture transformation using promoters and DNA constructs
XX PS Example 2: Column 41-44; 49pp; English.
XX CC This sequence corresponds to the tobacco (Nicotiana tabacum) anthranilate
XX CC synthase ASAI. The corresponding gene's promoter and DNA constructs
XX CC containing it are useful for imparting, to a plant cell, tolerance to
XX CC an amino acid analogue of Trp, or altering the Trp content in a plant
XX CC by transforming the plant cells with an expression cassette containing
XX CC the ASAI structural gene. The promoter and construct are useful for the
XX CC production of transgenic plants.
XX SO Sequence 491 AA;

Query Match 10.4%; Score 209; DB 20; Length 491;
Best Local Similarity 26.4%; Pred. No. 2e-11;
Matches 77; Conservative 56; Mismatches 123; Indels 36; Gaps 9;

QY 110 QFVPRQNTVTTSHTVDEVMWIERENLIDTLAIDOTLAKVFGROOTLSDTL 164
DB 191 qfprslhn-----snlce---eykeavvk--akehllagdlfqlvlsqrfertrfddpf 240
QY 165 RLAQIIRALAEQANFYHVLKRHDELFSATPERLVAMSGQIATAVAGSRGTCGAD 224
DB 241 eyrtrrlv---npsymyylgargcllvasspelltkkqkivnrplagctskrykneve 298
QY 225 DIALGELLASQKNRIEQYVASTITRLQDVT--SLKVPAMPSLKKNQVOHLYTPT 282
DB 299 dkrl-eellenekegaehlmvelgrndygvtkysvkeklmrleryshvmbhsstvt 357
QY 293 GDIAHLSTAVLDLHPTPLGVPREALYYIATHEKTPRGAGPIGFTTADNSEEFG 342
DB 358 gelqgqlctwdvlraalpvgvtsgapkxkamellleleptlrpgysggfgyvsftgdmndi 417
QY 343 VVGRISM-----YVNOTQR--ATLFAAGIYVADSDAQOEVEE 378
DB 418 alsrltvtvctacqyntmsykdankrtrewaylqagagvvaadsdpqdehce 469

RESULT 8
Y42112
ID Y42112 standard; Protein; 603 AA.
```

AC	Y42112,	
XX		
DT	13-DEC-1999	(first entry)
XX		
DE	Corn anthranilate synthase alpha subunit protein sequence.	
KW	Anthranilate synthase alpha subunit; tryptophan biosynthetic enzyme;	
KW	chromic gene; corn; rice; soybean; wheat; tryptophan synthase;	
KW	anthranilate synthase beta subunit; herbicide; fungicide; phenotype;	
KW	pathogen attack; identification; transgenic plant; ASAs; ASBs; TSAs;	
KW	plant breeding.	
XX		
OS	Zea mays.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 15	/note= "unspecified"
FT	Misc-difference 75	/note= "unspecified"
FT	Misc-difference 115	/note= "unspecified"
FT		/note= "unspecified"
XX		
PN	W09949058-A2.	
XX		
PD	30-SEP-1999.	
XX		
PF	19-MAR-1999; 99WO-US06046.	
XX		
PR	26-MAR-1998; 98US-0079386.	
XX		
PA	(DUPO ) DU PONT DE NEMOURS & CO E. I.	
XX		
PI	Vollmer SJ, Falco SC, Broglie RM, Bryan GT, Cahoon RE;	
PI	Rafalski JA;	
XX		
DR	WPI: 1999-580451/49.	
XX		
DR	N-PSDB; 225109.	
XX		
PT	New isolated tryptophan biosynthetic enzyme nucleic acids, used to	
PT	produce plants with altered tryptophan levels and for developing	
PT	herbicides or fungicides -	
XX		
PS	Claim 21; Page 55-57; 83pp; English.	
XX		
CC	The present invention describes isolated anthranilate synthase alpha	
CC	subunit (ASAs), anthranilate synthase beta-subunit (ASBs), and tryptophan	
CC	synthase alpha-subunit (TSAs) nucleic acids, and protein encoded by them,	
CC	obtained from corn, rice, soybean and wheat cDNA libraries. The nucleic	
CC	acid fragments may be used to create transgenic plants in which the	
CC	disclosed ASAs, ASBs or TSAs are present at higher or lower levels than	
CC	normal or in cell types or developmental stages in which they are not	
CC	normally found. This would have the effect of altering the level of	
CC	tryptophan in those cells. Manipulation of the levels of some of the	
CC	ASAs will also results in changes in the response to pathogen attack.	
CC	Because this pathway is not followed for the production of tryptophan in	
CC	higher animals, these enzymes are very good candidates for the discovery	
CC	of herbicides and fungicides. The ASAs, ASBs or TSAs can be used as	
CC	targets to facilitate design and/or identification of inhibitors of those	
CC	enzymes that may be useful as herbicides. Nucleic acid fragments can also	
CC	be used as probes for genetically and physically mapping the genes that	
CC	they are a part of, and as markers for traits linked to those genes. Such	
CC	information may be useful in plant breeding in order to develop lines	
CC	with desired phenotypes. 225109 to 225127 represent specifically claimed	
CC	nucleic acids from the present invention and Y42112 to Y42130 represent	
CC	the proteins encoded by them.	
XX		
XX		
XX	Sequence 603 AA:	
XX		

```

OY 168 Q1RL-LEQANTYHVVLKRRDELFIATPERLVAMSGQIATAVAGTSRRGTCDADI 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 eyvalitvimpsspymauyqaragcvlsspeflitvskgkrlmrlprlaglvtgktekdq 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 227 ALGELLASQKNRIEHOYVVAITTRLODYVT--SLVYPMAPSILSKKOVHLYTPTGD 284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 mgegqllddekgceehmlvdlgtndygvkskpsvkvcklmleturyshymbhlstcvsqg 470
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 285 IAAHLSVTAIVDRLHPRLALGCVPREAALYYIATNEKTPRGLEAGPICYFTADNSGEFVV 344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 lddhgwdsdalraalprvgctvgagkavkamelldklevtrrpyvsglfgdmqlal 530
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 345 G1RSM-----YVWQDOR---ATFLPAGACIVADSDAQOEYEE 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 slrtlvfstapshmtmvsykdadrtrewahlqgagvlvsspddeqre 560
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

SEQUENCE	9
ID	Y41301
AC	Y41301 standard; Protein; 616 AA.
CC	
CC	Y41301;
CC	18-JAN-2000 (first entry)
DE	Tobacco anthranilate synthase ASA2.
RW	Promoter; tobacco; anthranilate synthase; tolerance; analogue; Trp; tryptophan; transgenic plant.
OS	Nicotiana tabacum.
PN	US5965727-A.
PD	12-OCT-1999.
PF	31-DEC-1997; 97US-0001826.
PR	26-JUL-1996; 96US-0025140.
PR	25-JUL-1997; 97US-0937739.
PA	(UNII ) UNIV ILLINOIS FOUND.
PI	Widholm JM, Brotherton JE, Song H;
DR	WPI: 1999-579943/49.
XX	N-PDB; Z30444.
PT	Plant tissue culture transformation using promoters and DNA constructs
PS	-
XX	Example 2; Column 31-34; 49pp; English.
XX	This sequence corresponds to the tobacco (Nicotiana tabacum) anthranilate synthase ASA2. The corresponding gene's promoter and DNA constructs containing it are useful for imparting, to a plant cell, tolerance to an amino acid analogue of Trp, or altering the Trp content in a plant by transforming the plant cells with an expression cassette containing the ASA2 structural gene. The promoter and construct are useful for the production of transgenic plants.
XX	Sequence 616 AA;
XX	

Query Match	9.7%	Score 195	DB 20	length 616
Best local similarity	23.0%	Pred. No. 5	7e-10	
Matches	73	Conservative	69	Mismatches 141; Indels 34; Gaps 9
QY	84	EWAMT-----IAADKTTTSSDYSDTTWLAQFVPKQPNVTTSHTYDEVDMTERENLID	139	
Db	254	ELVSVRGVIGISPTSPSPS---vdctt--haqgp-----sltkgmntse-eyknnaviqake	303	
Y	140	TLAIDQTLAKVFGHQOQTLOLSDTLRLAQIIRALAEQANTYHVULKRHDELISATPERT	199	

[illegible]

```

RESULT      10
W93814
ID          W93814 standard; Protein; 606 AA.
AC
XX
AC          W93814;
XX
DT          25-JUN-1999 (first entry)
XX
DE          Rice anthranilate synthase second isozyme alpha-subunit protein.
XX
KW          Anthranilate synthase; alpha-subunit; ASA; rice; isozyme; plant; seed
KM          maize; wheat; tryptophan content; nutritional value.
XX
OS          Oryza sativa.
XX
PN          M099J1800-A1.
XX
PD          11-MAR-1999.
XX
PE          31-AUG-1998; 98WO-JP03883.
PR          29-AUG-1997; 97JP-0235049.
PA          (HOKK ) HOKKO CHEM IND CO LTD.
PI          (NORQ ) JAPAN MIN AGRIC FORESTRY & FISHERIES.
XX
PI          Hasegawa H, Terakawa T, Tozawa Y, Wakasa K;
XX
DR          WI: 1999-228982/19.
DR          N-PSDB; X23753.
XX
PT          DNA encodes ^a-subunit of first isozyme of rice anthranilate
PT          synthase - used for improving tryptophan production and
PT          nutritional value of crops, e.g. rice, maize or wheat
XX
PS          Claim 3; Page 139-141; 152pp; Japanese.
XX
CC          This invention describes a novel rice anthranilate synthase first
CC          isozyme alpha-subunit. The encoding DNA can be used to produce
CC          transformant plants and seeds, of e.g. rice, maize or wheat, with
CC          enhanced tryptophan content and nutritional value of the crops.
XX
SQ          Sequence      606 AA;

Query Match      9.4%; Score 189.5; DB 20; Length 606;
Best Local Similarity 26.7%; Pred. No. 1.8e+09;
Matches 62; Conservative % 45; Mismatches 102; Indels 23; Gaps

09      168 QIIRALAEQANTHYVUUKRDELFIISTPERLYAMSGGQATRAAVAGTSRRGTGDADDI 226
       1 : 111 : : : : : 1 : 1 : 111 111 : 1
       354 eyvriidilrvpsymayqagvcqlvvsspillrrvkylinplagtvrqktekede 413
```

```

Oy 227 ALSEALLASCKNRLEHQQVVASITRRLODYTT--SLKVPAMPBLKKKQVQHLPTED 264
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 mgeqqlIsdeqceahimlvoldgrndygvkskpsvsvvekImmleryshymhIsstvsge 473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 285 IAAHLSVTAIVDRLHPRPALGVGVREAAALVYIATHEKTPGRLFGPRLGYFTAAOISGEFVY 344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 IdhnhgswdaIraaIpygvtsygaqkvkamelIdelvtlrrgpyssgIqgIsfd--gdmlI 531
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 345 GI-----RSWY-VNORQR-----ATLPAGAGIVADSDAQOEYEE 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 alaIrtIvtscapshmtynsgYkdeIrtterrewahIqagagIvaospspddegre 583

```

RESULT	11
ID	Y42113
XX	Y42113 standard; Protein; 446 AA.
AC	
XX	Y42113;
XX	
DT	13-DEC-1999 (first entry)
XX	
DE	Corn anthranilate synthase alpha subunit portion protein sequence.
XX	
KW	Anthranilate synthase alpha subunit; tryptophan biosynthetic enzyme;
KW	chimeric gene; corn; rice; soybean; wheat; tryptophan synthase;
KW	anthranilate synthase beta subunit; herbicide; fungicide; phenotype;
KW	pathogen attack; identification; transgenic plant; ASAS; ASBS; TSAS;
KW	plant breeding.
XX	
OS	Zea mays.
XX	
FH	Key
FT	Misc-difference 31
FT	/note= "unspecified"
XX	
PN	W09949058-A2.
XX	
PD	30-SEP-1999.
XX	
PR	19-MAR-1999; 99WO-US06046.
XX	
PR	26-MAR-1998; 98US-0079386.
XX	
PA	(DUPO ) DU PONT DE NEMOURS & CO E I.
XX	
PI	Vollmer SJ, Falco SC, Broglie RM, Bryan GT, Cahoon RE;
PI	Rafalski JA;
XX	
DR	WPI; 1999-580451/49.
XX	
DR	N-PSDB; Z25110.
XX	
PT	New isolated tryptophan biosynthetic enzyme nucleic acids, used to
PT	produce plants with altered tryptophan levels and for developing
PT	herbicides or fungicides -
XX	
PS	Claim 21; Page 57-59; 83pp; English.
XX	
CC	The present invention describes isolated anthranilate synthase alpha-
CC	subunit (ASAS), anthranilate synthase beta-subunit (ASBS), and tryptophan
CC	synthase alpha-subunit (TSAS) nucleic acids, and protein encoded by them,
CC	obtained from corn, rice, soybean and wheat cDNA libraries. The nucleic
CC	acid fragments may be used to create transgenic plants in which the
CC	disclosed ASAS, ASBS or TSAS are present at higher or lower levels than
CC	normal or in cell types or developmental stages in which they are not
CC	normally found. This would have the effect of altering the level of
CC	tryptophan in those cells. Manipulation of the levels of some of the
CC	ASASs will also results in followed in the response to pathogen attack.
CC	Because this pathway is not followed for the production of tryptophan in
CC	higher animals, these enzymes are very good candidates for the discovery
CC	of herbicides and fungicides. The ASAS, ASBS or TSAS can be used as
CC	targets to facilitate design and/or identification of inhibitors of those
CC	enzymes that may be useful as herbicides. Nucleic acid fragments can also
CC	be used as probes for genetically and physically mapping the genes that





PT nutritional value of crops, e.g. rice, maize or wheat  
XX  
PS Claim 1; Page 128-131, 152pp; Japanese.  
XY

PS Claim 1; Page 128-131; 152pp; Japanese.  
 YX

CC This invention describes a novel rice anthranilate synthase first  
CC isozyme alpha-subunit. The encoding DNA can be used to produce  
CC transformatant plants and seeds, of e.g. rice, maize or wheat, with  
CC enhanced tryptophan content and nutritional value of the crops.  
XX  
SQ Sequence 577 AA.

CC enhanced tryptophan content and nutritional value of the crops.

CC transformant plants and seeds, of e.g. rice, maize or wheat, with enhanced trypsinogen content and nutritional value of the

CC Isozyme alpha-subunit. The encoding DNA can be used to produce  
CC transformant plants and seeds of a rice maize or wheat with

XX

Sequence 577 AA;

Query Match 9.0%; Score 181.5; DB 20; Length 577;  
Best Local Similarity 25.7%; Pred No 9.7e-09.

Best Local Similarity 25.7%; Pred. No. 9.7e-09;  
Matches 59; Conservations 45; Mismatches 107

Matches 59; Conservative 45; Mismatches 107; Indels 19; Gaps 4;

```

Oy      168  QIIRAL-AEQANTHVVLKRHDELFISATBERLVAAMSGGQIATPAANVGSTRBRGDGADDI 226
      326  evyraltivpyspmyajlgatgcllvasspelltrvkrctlvmpriagrltrkyskaed 385
Oy      227  ALGELLASOKNRIEHQVAVASITTRLDQYTT--SLKVPAMPSLKKNQOHLPTPTGD 284
      386  vleqltisdgqcaehimlvldgrndvgvkspgvsvkvellmveryshvmlstvtge 445
Oy      285  IAAHLSTAIYDRHPRPALGCVPREAALYYIATHEKTPGGLRGPICGTATDNSEGEVY 344
      446  ltrdlctwdaIraalprgvtvsqgkvraramldqmeqkmtgpyssgfgtvgvfrgmdmlal 505
Oy      345  GIRSM-----YVNOQPRR---ATLFAGAGIVADSAQOEYEE 378
      506  alrtivprgsrftctmvsydcnargswahlgagagivadsqppdeqne 555

```

Db 326 evyraltlvnpspymaylgargcllvasspelltrvektlvnrplagtlirrgkskaedk 385

QY 227 ALGELLASQKNRIEHQYVASITTRLQDVT--SLKVPAMPSLLKNQVQHLTPITGD 284

386 vleglllsdgqcaenhlmldgrndvgkvskspsvkvkvlmneverysmhvhlstvtge 445

265 LAHLSVIAIVDKLHPLEALGVPREALYXIA THEKTPRGLEAGPICGYFTADNSGEFV 344

[illegible][illegible]

Search completed: November 25, 2000, 02:17:00  
Job time: 3222 sec

Job time: 3222 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 01:48:48 ; Search time 41.98 Seconds  
(without alignments)  
158.509 Million cell updates/sec

Title: US-08-774-104A-2  
Perfect score: 2009  
Sequence: 1 MMTYHETPRLAQSDLOQLXA.....ETGLKEPRLQKDYHVE 397

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6.COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/6.COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/PCTUS.COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2009	100.0	397	2	US-08-774-104A-2
2	210.5	10.5	595	3	US-08-604-789B-3
3	209	10.4	491	2	US-09-001-826-23
4	201.5	10.0	621	3	US-08-604-789B-4
5	195	9.7	616	2	US-09-001-826-5
6	184.5	9.2	604	3	US-08-604-789B-16
7	183.5	9.1	604	3	US-08-604-789B-2
8	96.5	4.8	1346	3	US-09-320-878-4
9	94	4.7	703	4	PCT-US95-06994-8
10	94	4.7	3739	3	US-09-320-878-2
11	92.5	4.6	416	3	US-08-910-505-2
12	92.5	4.6	416	3	US-08-910-505-4
13	92.5	4.6	1562	3	US-09-320-878-3
14	92.5	4.6	3038	1	US-08-450-332-2
15	92.5	4.6	3038	2	US-08-637-640-2
16	91.5	4.6	4551	3	US-09-320-878-1
17	90.5	4.5	3031	1	US-07-689-008-2
18	90	4.5	636	1	US-08-765-081-5
19	90	4.5	636	1	US-09-098-082-5
20	90	4.5	636	4	PCT-US95-06994-5
21	88	4.4	1167	2	US-08-589-756-2
22	87.5	4.4	487	2	US-08-712-709-8
23	87.5	4.4	487	3	US-09-111-444-8
24	87.5	4.4	1026	1	US-08-194-290-7
25	87	4.3	718	4	PCT-US95-06994-6
26	87	4.3	887	4	US-08-215-709-1
27	86.5	4.3	787	1	US-08-574-763-2
28	85	4.2	800	1	US-08-785-052-4

29	85	4.2	800	2	US-08-913-581-4	Sequence 4, Appl1
30	84	4.2	878	1	US-08-237-919-2	Sequence 2, Appl1
31	84	4.2	878	4	PCT-US95-05518-2	Sequence 2, Appl1
32	83.5	4.2	339	1	US-08-433-854-4	Sequence 4, Appl1
33	83.5	4.2	339	1	US-08-174-745A-4	Sequence 4, Appl1
34	83.5	4.2	339	2	US-08-195-947-4	Sequence 4, Appl1
35	83.5	4.2	339	2	US-08-433-885-4	Sequence 4, Appl1
36	83.5	4.2	339	2	US-08-433-908B-4	Sequence 4, Appl1
37	83.5	4.2	7257	3	US-09-335-409-5	Sequence 5, Appl1
38	83	4.1	1150	2	US-08-589-756-3	Sequence 3, Appl1
39	82.5	4.1	529	1	US-08-548-509-2	Sequence 2, Appl1
40	82.5	4.1	1026	2	US-08-614-377A-7	Sequence 7, Appl1
41	82	4.1	1289	2	US-08-542-003-2	Sequence 2, Appl1
42	82	4.1	1289	2	US-08-322-760A-2	Sequence 2, Appl1
43	81.5	4.1	394	3	US-08-673-814-6	Sequence 6, Appl1
44	81.5	4.1	1164	2	US-08-589-756-1	Sequence 1, Appl1
45	81	4.0	2089	1	US-08-418-893D-23	Sequence 23, Appl1

## ALIGNMENTS

RESULT 1  
US-08-774-104A-2  
Sequence 2, Application US/08774104A  
Patent No. 5919687  
GENERAL INFORMATION:  
APPLICANT: Chatterjee, Subroto  
TITLE OF INVENTION: RECOMBINANT N-SMASES AND NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/774,104A  
CLASSIFICATION: 435  
FILING DATE: 12/24/96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Corleiss, Peter F  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 46906  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-774-104A-2  
Query Match 100.0%; Score 2009; DB 2; Length 397;

Best Local Similarity 100.0%; Pred. No. 9.2e-203;  
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MMTYETRALAOSDLOQYLALETTEFGAYFATPADDLRFICGATATKKTAAQLOGAVF 60
Db 1 MMTYETRALAOSDLOQYLALETTEFGAYFATPADDLRFICGATATKKTAAQLOGAVF 60
QY 61 GAOSFEDEQEPSELMAGFWFVPEVWVTIAADKITEGSDTVSDFTTWLAQFVKOPNTVT 120
Db 61 GAOSFEDEQEPSELMAGFWFVPEVWVTIAADKITEGSDTVSDFTTWLAQFVKOPNTVT 120
QY 121 TSHVTDVDMWIRTEMLIDTLAIDOTLAKVFGROQTLOLSPTLRALQIRALAEQANTY 180
Db 121 TSHVTDVDMWIRTEMLIDTLAIDOTLAKVFGROQTLOLSPTLRALQIRALAEQANTY 180
QY 181 HVLKRHDELFSATPERLYVAMSGGIATAVAGTSRRGTGADADTALCEALLASOKNRI 240
Db 181 HVLKRHDELFSATPERLYVAMSGGIATAVAGTSRRGTGADADTALCEALLASOKNRI 240
QY 241 EHQYVVASITTRLODYTTSLKVPAMPSSLKKNQOVHLYPTIGDIAHLSTVATVDRLRP 300
Db 241 EHQYVVASITTRLODYTTSLKVPAMPSSLKKNQOVHLYPTIGDIAHLSTVATVDRLRP 300
QY 301 TPALGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYVNOTORRATL 360
Db 301 TPALGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYVNOTORRATL 360
QY 361 FAGAGIVADSDAQOEYEETGLKFEPRKOLLKDYNHVE 397
Db 361 FAGAGIVADSDAQOEYEETGLKFEPRKOLLKDYNHVE 397
```

## RESULT 2

US-08-604-789B-3

Sequence 3, Application US/08604789B

Patent No. 6118047

GENERAL INFORMATION:

APPLICANT: Anderson, P.C.

Chomet, P.S.

Grifford, M.C.

Kilz, A.L.

TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE

AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN

OVERPRODUCTION

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESS: Schwegman, Lundberg, Woessner &amp; Kiluth, P. A.

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/604,789B

FILING DATE: 19-Jan-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: &lt;Unknown&gt;

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Woessner, Warren D.

REGISTRATION NUMBER: 30,440

REFERENCE/DOCKET NUMBER: 950.026US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (612) 373-6903

TELEFAX: (612) 339-3061

TELEX: &lt;Unknown&gt;

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 595 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-604-789B-3

## Query Match

10.5%; Score 210.5; DB 3; Length 595;

Best Local Similarity 26.0%; Pred. No. 1.7e-13;

Matches 76; Conservative 57; Mismatches 124; Indels 35; Gaps 8;

```
QY 110 QFVPKOPNTVTTSHVTDVDMWIRTEMLIDTLAIDOTLAKVVF-----GROQTLOLSPTL 164
Db 294 QFVPSLDN-----SNVTC-----EYKAVYK--AKEHILAGDIFQIYLSQRFERRTADPF 343
QY 165 RLAQIIRALAEQANTYHVLKRHDELFSATPERLYVAMSGGIATAVAGTSRRGTGAD 224
Db 344 EYVRAIRLV--NPSPYMGYLQARGCILVASSPEILTKVKNKIVNRPPLAGTSKRGKNEVE 401
QY 225 DIALGALLASOKNRIEHQYVVASITTRLODYTT--SLKVPAMPSSLKKNQOVHLYPTIT 282
Db 402 DKRLKELLENEKQCAEHMLVDLGRNDGKTYKGSVVEKLMNTERSHVHHISSTVT 461
QY 283 GDIAHLSTVATVDRLRHPTPALGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEF 342
Db 462 GELQSDLTQMDVIRALRPYGTSGAPKAVKAMELIDLEFTRRCPYSGFGGVSFTDDMDI 521
QY 343 VVGIRSM-----YVNOTOR--ATLFAGAGIVADSDAQOEYEE 378
Db 522 ALSRTIVFPACQYNTMYSKDANKRRVAVYLQAGAGVADSDPDQDCE 573
```

## RESULT 3

US-09-001-826-23

Sequence 23, Application US/09001826A

Patent No. 5965727

GENERAL INFORMATION:

APPLICANT: SONG, HEE-SOOK

APPLICANT: BROUGHTON, JEFFREY E.

APPLICANT: WIDHOLM, JACK M.

TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE

FILE REFERENCE: U1001.C1

CURRENT APPLICATION NUMBER: US/09/001,826A

CURRENT FILING DATE: 1997-12-31

EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140

EARLIER FILING DATE: 1997-07-25; 1996-07-26

NUMBER OF SEQ ID NOS: 31

SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text ;

SEQ ID NO 23

LENGTH: 491

TYPE: PRT

ORGANISM: Nicotiana tabacum

US-09-001-826-23

Query Match 10.4%; Score 209; DB 2; Length 491;  
Best Local Similarity 26.4%; Pred. No. 1.8e-13;  
Matches 77; Conservative 56; Mismatches 123; Indels 36; Gaps 9;

```
QY 110 QFVPKOPNTVTTSHVTDVDMWIRTEMLIDTLAIDOTLAKVVF-----GROQTLOLSPTL 164
Db 191 QFVPSLDN-----SNVTC-----EYKAVYK--AKEHILAGDIFQIYLSQRFERRTADPF 240
QY 165 RLAQIIRALAEQANTYHVLKRHDELFSATPERLYVAMSGGIATAVAGTSRRGTGAD 224
Db 241 EYVRAIRLV--NPSPYMGYLQARGCILVASSPEILTKVKNKIVNRPPLAGTSKRGKNEVE 298
QY 225 DIALGALLASOKNRIEHQYVVASITTRLODYTT--SLKVPAMPSSLKKNQOVHLYPTIT 282
Db 299 DKRL-EELLENEKQCAEHMLVDLGRNDGKTYKGSVVEKLMNTERSHVHHISSTVT 357
```

QY	283	GDIAAH:SVIAIVRLHPALGVPREALYYIATHEKPRGLFAGPIGYFADNSGEE	342
Db	358	GELODGLFCMDVLAALPVGVSAPKVKAMELIDELFPRRGPYSGFGVSGFTGMDI	417
QY	343	VGIRSM-----VYNOTOR---ATLFAGAGIVADSDAQOEYEE	378
Db	418	ALSTRIVFPACQNTMYSKYDANKRREWVAYIQAGAGVADSDPDDEHCE	469
RESULT 4			
		US-08-604-789B-4	
		Sequence 4, Application US/08604789B	
		Patent No. 6118047	
		GENERAL INFORMATION:	
		APPLICANT: Anderson, P.C.	
		Chomet, P.S.	
		Griffor, M.C.	
		Krlz, A.L.	
		TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE	
		AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN	
		OVERPRODUCTION	
		NUMBER OF SEQUENCES: 16	
		CORRESPONDENCE ADDRESS:	
		ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.	
		STREET: P.O. Box 2938	
		City: Minneapolis	
		STATE: MN	
		COUNTRY: USA	
		ZIP: 55402	
		COMPUTER READABLE FORM:	
		MEDIUM TYPE: Diskette	
		COMPUTER: IBM Compatible	
		OPERATING SYSTEM: DOS	
		SOFTWARE: FastSeq Version 2.0	
		CURRENT APPLICATION DATA:	
		APPLICATION NUMBER: US/08/604,789B	
		FILING DATE: 19-Jan-1996	
		PRIOR APPLICATION DATA:	
		APPLICATION NUMBER: <Unknown>	
		FILING DATE: <Unknown>	
		ATTORNEY/AGENT INFORMATION:	
		NAME: Woessner, Warren D.	
		REGISTRATION NUMBER: 30,440	
		REFERENCE/DOCKET NUMBER: 950,026US1	
		TELECOMMUNICATION INFORMATION:	
		TELEPHONE: (612) 373-6803	
		TELEFAX: (612) 339-3061	
		TELEX: <Unknown>	
		INFORMATION FOR SEQ ID NO: 4:	
		SEQUENCE CHARACTERISTICS:	
		LENGTH: 621 amino acids	
		TYPE: amino acid	
		STRANDEDNESS: single	
		TOPOLOGY: linear	
		MOLECULE TYPE: protein	
		SEQUENCE DESCRIPTION: SEQ ID NO: 4:	
		US-08-604-789B-4	
QY		Query Match	10.08; Score 201.5; DB 3; Length 621;
Db		Best Local Similarly	24.08; Prid. No. 1.6e-12;
		Matches 69; Conservative	56; Mismatches 134; Indels 29; Gaps 6;
QY	111	FVPPQNPVTSTSHTDVDMIERENTLI--DLAIDOTLAKVYFGRQOTQLSDTLRLAQ	168
Db	308	FGPKLESTYMTSEAYKKA-VVEAKEHLADIDPI-----VLSQREKRTFADPEIYR	360
QY	169	IIRLALDAQANTYHVVLKRHDELFIATPERLVAMSGQIATAAVAGTSRRGTGAGADIAL	228
Db	361	ALRIV--NPSPYMAVLQVRCILVASSPELLLSKNNKINRPLAGTVRRGKTPKEDIMI	418
QY	229	GEALLASOKNRIEQIYVVASITTRLQDVT--SLKYPAMPSLKNNQOVHLVYPTTGIDIA	286

```

Db      419  EKLLSDKQCAEHIMLMDLGRNDGVKSPSCSVEVKKLKLDIEMFSHVMISSYVGGELL 478
Oy      287  AHLSTALVDRLRHPALGVPREALYIATHEKTPRCLFPGPIGYFADNSGEFVVI 346
Db      479  DHLTSMDLARALPVGTSGAPKAKAMELIDELVTRRQPYSGGFGGISFNDMDIAL 538
Oy      347  RSMYVNGQORATLF-----AGAGIVADSDAQOEYEE 378
Db      539  RTMVEPTTRDYDTLXSYKHPQRREMIATQAGAGIVADSNPDDEHRE 586

RESULT      5
US-09-001-826-5
; Sequence 5, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
; APPLICANT: SONG, HEE-SOOK
; APPLICANT: BROTHERTON, JEFFREY E.
; APPLICANT: WIDHOLM, JACK M.
; TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
; TITLE OF INVENTION: TRANSFORMATION
; FILE REFERENCE: UI001.C1
; CURRENT APPLICATION NUMBER: US/09/001,826A
; CURRENT FILING DATE: 1997-12-31
; EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140
; EARLIER FILING DATE: 1997-07-25; 1996-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: MacIntosh WordPerfect converted to PC ASCII Text
; SEQ ID NO 5
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-09-001-826-5

Query Match      9.7%; Score 195; DB 2; Length 616;
Best Local Similarity 23.0%; Pred. No.7.6e-12;
Matches 73; Conservative 69; Mismatches 141; Indels 34; Gaps 9

Oy      84  EVMVY-----IADKITTGSDYVSDEFTTWLAQVVPKQPNVTTSHTVDEVDWIERENTLID 139
Db      254  ELVSRVQGIESPRLSPGS---VDICT--HAGCP-----SLTGNNMTSE-EYKNAVLOAKE 303
Oy     140  TLAIQDTLAKVVGSRQQTQLQSLDTRLAQIIRALAEQANTYHVVLKRHDELFI SATPERL 199
Db     304  HIAADIFQIVLQSFERRTFADPREVYRALRIY--NPSPYMTYIQA RCILIVASSPEIL 361
Oy     200  VAMSGGQIATAVAATSRGTGADGDIALGEALLSOKKRIEHQYVVASITTRLQDYTT- 258
Db     352  TRVKRRRIYVNRPLAITSRGKTPDEBDVMEQMLDDEKRAEHIMLVJDIGRNDVGRVSKRP 421
Oy     259  -SLKVPAPSLILKNQVOYOLVTPITGDIANHLVYTAIVDRLRHPTRPALGVPREALYIYA 317
Db     422  GSVNVEKILMSVRYSHVMIHISTVSGELLDHITCMDALALPVGTVSAPVYKAMELID 481
Oy     318  THEKTPRCLFAGPIGYFADNSGEFVVGIRSM-----VYNDQORR---ATLF 361
Db     482  QLEVARRGPRYSGFGGISPSGDMIDIALALRTWVPLNGARYDTMYSTASKRQEWVAHLQ 541
Oy     362  AGAGIVADSDAQOEYEE 378
Db     542  SGAGIVADSNPDDEQIE 558

RESULT      5
US-08-604-789B-16
; Sequence 16, Application US/08604789B
; Patent No. 6118047
; GENERAL INFORMATION:
; APPLICANT: Anderson, P.C.
; Chomet, P.S.
; Gilfor, M.C.

```

Kriz, A.L.  
TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE  
AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN  
OVERPRODUCTION  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.  
STREET: P.O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/604,789B  
FILING DATE: 19-Jan-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Woessner, Warren D.  
REGISTRATION NUMBER: 30,440  
REFERENCE/DOCKET NUMBER: 950.0260S1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (612) 373-6903  
TELEFAX: (612) 339-3061  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 16:  
US-08-604-789B-16

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

[illegible]

```

: CURRENT APPLICATION NUMBER: US/09/320, 878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119, 139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100, 880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087, 080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1346
: TYPE: PRT
: ORGANISM: Streptomyces venezuelae
US-09-320-878-4
```

```

Query Match          4.8%; Score 96.5; DB 3; Length 1346;
Best Local Similarity 24.2%; Pred. No. 0.62;
Matches 73; Conservative 30; Mismatches 118; Indels 81; Gaps 15;
```

```

QY 3 TYHETRALAOSDLOO-LYALETTEFGAYFPADDTLRFSGIATAPAKTAQALOGAVEG 61
D 523 TAMEHRAVAVGDSREALDALRME-GIVRGTVDP-----GRVAFVFPQGTQWAGM 575
QY 62 AQSDEGEYPOSELMAGWFPEVFWVTIAADKITFGSDTVSDFTTWLAQFPKQPNVTY 121
D 576 AELIDSS-----PEFAAACE-----TALSPYDWSLEAVRQAPSAPT 616
QY 122 SHVTDEVMIERTENLIDTLAIDQTLAKV-----VGRQO-----TLQLS 161
D 617 ---LDRDVYQPV-----TFVWVSLAKVQNHGHTPEAVIGHSGEIAAAYVAGALFLD 668
QY 162 D-----TLRLAQITRALAEQANTYHVVLKRHDELFISATPERLYAMSGQIATAVAGTS 216
D 669 DAARVTVLRKSIAMHLAGKGMISLASE-----EATRQRIENLHG--LSIAAVNGPT 720
QY 217 RRGTDGADDLALGALLASQKNRI-----EHQYVASITTRLODYTTSL--KVPA 264
D 721 ATTVSG-DPQIQDLAOCADGIRARIIPYDASHSAHVEIENELADVLAGLSPQTPQ 779
QY 265 MP 266
D 780 VP 781
```

```

RESULT 9
PCT-US95-06994-8
: Sequence 8, Application PC/TUS9506994
: GENERAL INFORMATION:
: APPLICANT: Children's Hospital & Medical Center
: APPLICANT: University of Washington
: APPLICANT: Washington State University Research Foundation
: APPLICANT: TARR, PHILIP I
: APPLICANT: BILGE, SIMA S
: APPLICANT: BESSER, THOMAS E
: APPLICANT: VART JR, JAMES C
: TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPILIC
: STREET: SUITE 2800, 1420 FIFTH AVENUE
: CITY: SEATTLE
: STATE: WASHINGTON
: COUNTRY: USA
: ZIP: WA 98101
: COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06994
: FILING DATE: 07-JUN-95
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/265, 714
: FILING DATE: 24-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BRODERICK, THOMAS F
: REGISTRATION NUMBER: 31,332
: REFERENCE/DOCKET NUMBER: CHOR-18591
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 682 8100
: TELEFAX: (206) 224 0779
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 703 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
: DESCRIPTION: wherein "Xaa" residues represent gaps
: DESCRIPTION: introduced to facilitate best alignment with
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli O157:H7
: STRAIN: 86-24 NAHR
PCT-US95-06994-8
```

```

Query Match          4.7%; Score 94; DB 4; Length 703;
Best Local Similarity 22.5%; Pred. No. 0.39;
Matches 76; Conservative 37; Mismatches 125; Indels 100; Gaps 16;
```

```

QY 63 QSFDEQYPOSELMAGWFPEVFWVTIAADKITFGSDTVSDFTTWLAQFPKQPNVTTS 122
D 265 QRYDNRDQGLSLNGY-----DRLRYERNKTSAGYDGTFTFGTWKSY----- 308
QY 123 HVTDEVMIERTENLIDTLAIDQTLAKVFGROQTLQSDTLRLAQITRALAEQANTYHV 182
D 309 -----LNMNE--TENK-----GRELVRSVLKR-----DKWGLAGOPRELKES----NL 345
QY 183 VLKRHDELFISATPERLYAMSGQ-----INTAAVAGTSRRGTD-----GADDIALG 229
D 346 IL--NSLLTTPLGESHLYVVGGEFQSSMKDGVVLASTGETFEROKSKWSVFAEDEMHLT 402
QY 230 EALLASQKNRIHQ-----YVASITTRLODYTTSLKVPAMPISLKNKOYQ 275
D 403 DALATAGSREYHEHQFGHFSRAYIWDVADAMTLKGVTGYNKAPRMQGLHNG----- 458
QY 276 HLYTPTGDIAHLSVTALVDRLHPTPALG---VPREALYYITTHEKTPRGFLFAGPIG 332
D 459 -----ISG-----VSGOGKTNLGNDELPEBSVSYEAGVYVDNPAGLMANVTG 502
QY 333 YFTADNNGEFVYGI-----RSMYVNGQTOR--ATLFAG 363
D 503 FWT-DESNKIXYSINDNTNSIVNSGARLHGVFEAG 539
```

```

RESULT 10
US-09-320-878-2
: Sequence 2, Application US/09320878A
: Patent No. 611/659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
```

```
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 3739
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-320-878-2
```

Query Match 4.7%; Score 94; DB 3; Length 3739;

Best Local Similarity 22.4%; Pred. No. 6.1; Mismatches 143; Indels 128; Gaps 20;

```
Matches 89; Conservative 36;
QY 6 ETRALA-----QSDLOQLYALLETTERGAYFATPADTLRFGIGALITAKTAQALOGAVFG 61
DB 2033 EHRAVALGTGODDLAALALAPBGLVRG-----VASGVRAVAFVPGGTQWAGMG 2082
QY 62 AGSFDEQETPQSLMAGFMFVPEVMTIADKTTFGSDTVSDFTYLAQVPRQPTVTY 121
DB 2083 AELIDVSK-----EFAAMMECEA-----ALAPYVDSLEAVRQAPCAPT 2123
QY 122 SHVTDEVDWIERTENLIDTLAIDOTLAKV-----VFGRQO-----TLQLS 161
DB 2124 LERVNDVGVY-----TFAVVSLAKVQNHGYPPOAVVGHSGGEIAAAYAGALSLD 2175
QY 162 D-----TLRLAQIIRALAQANTYHVLKRHDELFTSAPPERLVAMSGOIATPAVAGTS 216
DB 2176 DAAVYVTLKRSKSGIHLAQGGLSLASE-----AAVERLAGPDG--LSVAAYVNGPT 2227
QY 217 RRGTDGADIDALGELLASOKNRI-----EHQYVASTITTLQDYTTSL--KVPA 264
DB 2228 ATVVSQ-DPTQIOELAQACEADGVRRARIIPVDYASHSAHVETTESLADVLAGLSPQTPQ 2286
QY 265 MP--SLKKN-----KOVQHL--YTPITGDIANLSVTAIVD-RLHP--TPA 303
DB 2287 VPFESTLEGAWITERPALDGYWRNLRHKGVPAPVETLATIDGFTHFPEVSAHPVLTMA 2346
QY 304 LCGVPREALYYIATHEKTPRGILFAGPIGYFTADNSGE 341
DB 2347 L-----PEYVTGL-----GTLRDRNGGQ 2364
```

```
RESULT 11
US-08-910-505-2
Sequence 2, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
```

```
APPLICANT: Lonetto, Michael
APPLICANT: Warren, Patrick
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 416
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-08-910-505-2
```

Query Match 4.6%; Score 92.5; DB 3; Length 416;

Best Local Similarity 20.1%; Pred. No. 0.23; Mismatches 154; Indels 99; Gaps 20;

```
Matches 80; Conservative 66;
QY 5 HETRALAQSDLOQLYALLETTERGAYFATPADTLRFGIGALITAKTAQALOGAVFGAQS 64
DB 35 HGDKALKMYNLTPDHTTDLHLSHEQIKAAFDTL-----DETKQALQOSYERIK 86
QY 65 FDE-----QEPQS-ELMAGFMFVPEVMTIADKTTFGSDTVSDFTYLAQ----- 110
DB 87 YQESIKOTNOQLSESVCEIYHPLBSGIVYVPGKASYSYVL--MTATLAQAVGENT 144
QY 111 --FVPRQPTVT-----TSHTVDEVDWIERTENLIDTLAIDOTLAKVPRQOTLQSDPT 163
DB 145 VVTPPQPNQVSGEVLAACTITG-----VQVQVQVGAQSIATLTVG--TETIPKVD- 194
QY 164 LRLAQIIRALAQANTYHVLKRHDELFTSAPPERLVAMSGOIATPAVAGTSRSG--T 220
DB 195 -----KIVGRNQGFVAATAKTY--LF-----GQVGDQIAGPTEIALIID 231
QY 221 DGAD-DIALGELLASOKNRIEHQYVASTITTLQDYTTSLKVPAMSL--LKNQV- 274
DB 232 DTADLDIAIVYDFAQAEHDELARTYVIGEDAQVLKLESRI-AKALPNVDRYDIVSKSIA 290
QY 275 -QH-LVTPITGDIANLSV-----AIVDLRHPTRPALGVPREALYYIATHEKTPRGILF 327
DB 291 NQHYLLHASNFDRCCHVMNTIAPBHASIQVNPQPIYEKKYVGALE----- 337
QY 328 AGPIGYFTADNSGEFVVGIRSMYVNCORRATFPAGAGI 366
DB 338 ---IGHSPVIGDIYVAG--PSHYLPTNKRTARTNGLSV 371
```

```
RESULT 12
US-08-910-505-4
Sequence 4, Application US/08910505A
Patent No. 6107071
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Raymond
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
APPLICANT: Lonetto, Michael
APPLICANT: Warren, Patrick
TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
FILE REFERENCE: P50549-02
CURRENT APPLICATION NUMBER: US/08/910,505A
CURRENT FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 416
TYPE: PRT
```

```

RESULT 13 378
US-09-320-878-3
; Sequence 3, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320, 878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119, 139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100, 880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087, 080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1562
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-320-878-3
Query Match 4.6%; Score 92.5; DB 3; Length 1562

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```







**THIS PAGE BLANK (USPTO)**

Mon Dec 4 10:44:32 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 01:55:53 ; Search time 51.06 Seconds  
(without alignments)  
493.414 Million cell updates/sec

Title: US-08-774-104a-2

Perfect score: 2009  
Sequence: 1 MMTYHETRALAQSDQLQLYA.....ETGLKFEPMRQLKDYNHVE 397

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR65:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.5	24.7	471	1	A69657 probable isochoris
2	453	22.5	476	1	S58229 pcha protein - pse
3	383	19.1	474	1	S75568 isochorismate synt
4	354.5	17.6	372	2	H70595 probable entc prot
5	344	17.1	391	1	SYECIX isochorismate synt
6	328.5	16.4	398	1	C69615 isochorismate synt
7	326.5	16.3	396	1	A40365 siderophore biosyn
8	317.5	15.8	356	2	G64997 isochorismate synt
9	289	14.4	421	2	JC5323 anthranilate synth
10	278.5	13.9	523	2	A42301 anthranilate synth
11	278	13.8	430	2	E64059 probable isochoris
12	272.5	13.6	456	2	S35124 anthranilate synth
13	257	12.8	516	2	G70556 probable trpe prot
14	253	12.6	511	2	T36306 probable anthranil
15	252.5	12.6	433	2	T43924 anthranilate synth
16	251.5	12.5	434	2	A75163 anthranilate synth
17	249.5	12.4	470	2	A37854 para-aminobenzoate
18	246.5	12.3	494	2	E70352 anthranilate synth
19	246	12.2	464	2	C69088 anthranilate synth
20	245	12.2	461	2	D72414 anthranilate synth
21	245	12.2	515	1	JNBS1 anthranilate synth
22	244.5	12.2	512	2	JH0098 anthranilate synth
23	244	12.1	462	2	A40362 anthranilate synth
24	236	11.7	529	2	T45254 probable anthranil
25	234.5	11.7	317	2	B72489 probable anthranil
26	232	11.5	474	2	B64434 anthranilate synth
27	227	11.3	502	2	T35072 probable anthranil
28	223.5	11.1	416	2	H81376 probable anthranil
29	220	11.0	497	2	S11891 anthranilate synth

30	219	10.9	508	2	S74537 anthranilate synth
31	218.5	10.9	489	2	T40974 anthranilate synth
32	217.5	10.8	454	2	A31132 p-aminobenzoate sy
33	217	10.8	494	2	JX0065 anthranilate synth
34	217	10.8	508	2	S19266 anthranilate synth
35	216.5	10.8	434	2	T17434 anthranilate synth
36	215.5	10.7	638	2	F75547 anthranilate synth
37	215	10.7	485	2	S75655 anthranilate synth
38	215	10.7	503	2	T46852 anthranilate synth
39	214.5	10.7	473	2	E75353 anthranilate synth
40	213.5	10.6	411	2	B69450 anthranilate synth
41	213.5	10.6	507	1	NNBY1 anthranilate synth
42	211.5	10.5	491	2	E81132 anthranilate synth
43	210.5	10.5	364	2	T43181 probable anthranil
44	210.5	10.5	491	2	G81892 probable anthranil
45	210.5	10.5	595	2	J01684 anthranilate synth

## ALIGNMENTS

RESULT 1  
A69657  
probable isochorismate synthase (EC 5.4.99.6) menaquinone-specific menF - Bacillus su  
C:Species: Bacillus subtilis  
C>Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 16-Jun-2000  
C/Accession: A69657; S27507; S27508; I39883; T46639; T46639  
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fuma, S.; Gallizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M  
Koelter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masudaj, S.; Mau  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toynoni, K.; Yasumoto, K.; Yoshida  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; M01D:98044033  
A:Accession: A69657  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-471 <KUN>  
A:Cross-references: GB:299119; GB:AL009126; NID:g2635411; PIDN:CAB15061.1; PID:g26355  
A:Experimental source: strain 168  
R:Rowland, B.; Hill, K.; Mueller, J.; Driscoll, J.; Taber, H.  
Submitted to the EMBL Data Library, October 1991  
A:Description: Organization of an operon involved in menaquinone biosynthesis in Bact  
A:Reference number: S27507  
A:Accession: S27507  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', 13-110, 'WDLCYSEDPLTLAKKVNHTSRKGLSLCLRLC' <ROW1>  
A:Cross-references: EMBL:M74538; NID:g1185287  
A:Accession: S27508  
A:Molecule type: DNA  
A:Residues: 151-471 <ROW2>  
A:Cross-references: EMBL:M74538; NID:g1185287  
R:Miller, P.  
J. Bacteriol. 170, 2742-2748, 1988  
A:Title: Transcriptional regulation of a promoter in the men gene cluster of Bacillus  
A:Reference number: I39883; M01D:88227858  
A:Accession: I39883  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', 13-32 <MTL>  
A:Cross-references: GB:M21320; NID:g143178; PIDN:AAA22594.1; PID:g551715  
R:Driscoll, J.R.; Taber, H.W.  
J. Bacteriol. 174, 5063-5071, 1992  
A:Title: Sequence organization and regulation of the Bacillus subtilis menBE operon.

A:Reference number: A42715; MUID:92332443  
A:Accession: T46638  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-11, 'K', '13-110, 134-135, 'LCYSE', '141-142, 'LT', 'T', '317, 'ANKEVHN', '325, 'T', '327, 'A  
A:Cross-references: EMBL:M74521; NID:g557486; PIDN:MAA50396.1; PID:g557487  
A:Accession: T46639  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 151-471 <DR2>  
A:Cross-references: EMBL:M74521; NID:g557486; PIDN:MAA50397.1; PID:g557488  
A:Genetics:  
A:Gene: menF  
C:Superfamily: isochorismate synthase  
C:Keywords: Intramolecular transferase; isomerase

Query Match	24.78;	Score 496.5;	DB 1;	Length 471;
Best Local Similarity	-30.28;	Pred. No. 2.4e-30;		
Matches 127;	Conservative 77;	Mismatches 161;	Indels 55;	Gaps 10;

OY	19	YAALETTFFGAVYFNRPADDTLRFGLGATATKTKQO-----	53
Dd	46	YGAKKTYTNRFFWMSDBESELITVGLGKFAVQTQNKSSERKREVFEDOMERKKAFTAHITE	105
OY	54	-----ALGAVFAGSOFDEOEYPOSE--LMAGFWFVEPVMTIADK-----	95
Dd	106	EKKLDHNSAVGBVLTFGGFSGDCEBRSQMDHFSGDFEFLVAMLTMTAEGPFLTVNRWS	165
OY	96	FGSD---TVSDPTTWLAQF-VP--KQRP--IVTTSHTYDEVMLTERENLIDTLAIDQTL	147
Dd	166	GGEBAEAVLEGLKAFPAAEFWPDERKOEQAVIAAAEELDKDMKALKEGTASQIK--EKQY	224
OY	148	AKVVFROOTQLOSDTLAOLIIIALA--EQAMTIVHYVLKRHDELFIATPERLAVMSGQ	206
Dd	225	DKVYLARLLLTFBGPPIQIEPVLTLLDDDOQTSTVFPIEDQCKTFFGASPBRLKKGQGT	284
OY	207	IATAVACTSRRGTDGADDIALLGECALLASOKNRIEHOYVVASTITRLODYVTSLSKYPAMP	266
Dd	285	VMSGLGASIKRGVNEEDBRIIGLELLENKENLLEHDIIVGMIINAVSSCSSEVEKEDGP	344
OY	267	SLKNKQOVQHLTYPTIGDIAAHSVTAIVDRLHPALGCVPREALYIATHEKTPRGTL	326
Dd	345	VLYKTSVQHLFTPIVQGLRESASIFDLIEKTHPTPALGSGPOEKAVDVIRETEPMSRGW	404
OY	327	FAGRPCTYTAANSGSEFPVVCISMTNVNQTRAKTFLFAGACIYAASDAQOEVEBGLKREPM	386
Dd	405	YAAPGWMDSDQNEFEAFVAINSGILEGSTAR--FLFACGIVDESEPISEVEEDQIKTKPM	462

RESULT 2  
S58229  
pcha protein - *Pseudomonas aeruginosa*  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999  
C:Accession: S60203; S58229  
R:Serino, L.; Reilmann, C.; Baur, H.; Beyeler, M.; Visca, P.; Haas, D.  
Mol. Gen. Genet. 249, 217-228, 1995  
A:Title: Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*  
A:Reference number: S60202; MUID:96086939  
A:Accession: S60203  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-476 <SEZ>  
A:Cross-references: EMBL:X82644; NID:g1628425; PIDD:CA57969.1; PID:g929780  
A:Note: this protein is involved in salicylate biosynthesis  
A:Superfamily: Isochorismate synthase

Query Match	22.5%	Score 453;	DB 1;	Length 476;
Best Local Similarity	42.4%	Pred. NO. 5.2e-27;		
Matches 103; Conservative	38;	Mismatches 98;	Indels 4;	Gaps 3;

QY 169 KVVGRROOTLOLSDTLRLAQIIRNL-AEQANTYHVLLKKRHDELFIISTPERLVAMSGOI 207  
 Db 221 LVLIARTDARPLG-D-IEPMOYIEHLRIQHDAAQLFACRKNACGLGSPERLVRIRGEA 279  
 QY 208 ATAAVAGTSRRGTGADDDIALGELLASOKRIEHOYVVASITTRLODVTSLSKVPAMS 267  
 Db 280 LTHALAGTIARGGDAQEDARLGALLDSADKRNHEQLVNEAIRTALEPFSEVEIIPDAPC 339  
 QY 268 LKKKNVOVHETPLTGDIAHLASTALVDRHLHPRLAGGVPREALVYXITHEKTRGFL 327  
 Db 340 LKRLARVOHLETPRLRARADAGLIRLLQDLHPRLPAGGI-PRSNALDYIQHGMGRGW 399  
 QY 338 AGPIGYFTADNSGEFVYVGIIRSMVYQTORRATLFGAGIYADDAQOEYDEETGLKEPMR 387  
 Db 400 AALPGMLDGEENGDFLVALRSFALL--TPGRGIVYFAGCGIJDSEPAHEYRETCLIKLSAMR 457  
 QY 388 QLL 390  
 Db 458 EAL 460

RESULT 3  
575568  
isochlorismate synthase (EC 5.4.99.6) - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein slr0817  
C:Species: *Synechocystis* sp.

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C;Accession: s75568  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O.; K.; Okumura, S.; Shilimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu  
DNA Res. 3, 109-136, 1996

S: *Staphylococcus aureus* strain 8006-4  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S75568  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-474 <K&N>  
 A:Cross-references: EMBL:D90911; GB:AB001339; MIM:91653083; PIR:BA18129.1; PIR:q1655  
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Genetics:  
 A:Gene: entC  
 C:Superfamily: isochorismate synthase  
 C:Keywords: intramolecular transferase; isomerase

Query Match	19.1%;	Score 383;	DB 1;	Length 474;
Best Local Similarity	31.5%;	Pred. No. 1.2e-21;		
Matches 112;	Conservative 66;	Mismatches 150;	Indels 28;	Gaps 9;

[illegible]

Db 421 SEGNAEIVGIRSLAL--SRNRRLVYAGAGIVAGSDPLKEVAIEELKQLTMRSLI 474

RESULT 4

H70595 Probable entc protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70595

R:COLE, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 397, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:9825987

A:Accession: H70595

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-372 <COL>

A:Cross-references: GB:295120; GB:AL123456; NID:g3261739; PIDN:CAB08301.1; PID:g2072670

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: entC

C:Superfamily: Isochorismate synthase

Query Match 17.6%; Score 354.5; DB 2; Length 372;

Best Local Similarity 36.6%; Pred. No. 1.2e-19;

Matches 102; Conservative 35; Mismatches 129; Indels 13; Gaps 6;

Db 85 PKRVAAALPPADYLRIGRANDLAAFGDPLKVVLAAYQITADAPDARVLRV 144

QY 175 EQANTYHVVV-----KRHDE---LFTSATPERLVAMSGOIATAAVAGTSRGTGADDI 226

Db 145 VADPTAVGYLVDTLSAGNDGTGALVAGSPELLVANSNGNVKCFPGSAPRADAPRLDA 204

QY 227 ALGELLASQKNRIHQYVVAITRLQDVTSLSKVPAMSLKKNQVQHLVPTIGDIA 286

Db 205 ANAAALASSAKNRHEHDLVDTMKVLALEPLCEPLTTPAQPOLNKTAAVWHICTAITGRL- 263

QY 287 AHSVTAI--VDRLHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 344

Db 264 RNTSTTALDIALALHPPPAGVPTKATLELLELE-GDRGFYAGAVGCDGSGDGMVY 322

QY 345 GIRSTMYNOTORRATLFPAGAGIVADSDAQOEYEETGLKF 383

Db 323 SIRCAQLSADRRAALAHAGGIVAESDPDELETTTKF 361

RESULT 5

SYCIK Isochorismate synthase (EC 5.4.99.6) [validated] - Escherichia coli

N:Alternate names: isochorismate synthetase

C:Species: Escherichia coli

C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 20-Apr-2000

C:Accession: J0497; S04323; I73520; G64792

R:Ozenberger, B.A.; Brickman, T.J.; McIntosh, M.A.

J. Bacteriol. 171, 775-783, 1989

A:Title: Nucleotide sequence of Escherichia coli isochorismate synthetase gene entc and

A:Reference number: J0497; MUID:89123153

A:Accession: J0497

A:Molecule type: DNA

A:Residues: 1-391 <OZE>

A:Cross-references: GB:M24142; NID:g341117; PID:g450376

R:Elkins, M.F.; Earhart, C.F.

FEMS Microbiol. Lett. 56, 35-40, 1988

A:Title: An Escherichia coli enterobactin cluster gene with sequence homology to type an

A:Reference number: S04323; MUID:90236256

A:Accession: S04323

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304, 'TA', 307-391 <EIK>

A:Cross-references: GB:M36700; EMBL:X12670; NID:g145839; PIDN:AAA18491.1; PID:g145840

R:Brickman, T.J.; Ozenberger, B.A.; McIntosh, M.A.

J. Mol. Biol. 212, 669-682, 1990

A:Title: Regulation of divergent transcription from the iron-responsive fepB-entc pro

A:Reference number: 156426; MUID:90230305

A:Accession: 173520

A:Status: preliminary; translated from GB/EMBL/DD84

A:Molecule type: DNA

A:Residues: 1-33 <RES>

A:Cross-references: EMBL:X53274; NID:g48747; PIDN:CAA37371.1; PID:g48749

R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64792

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <BLAT>

A:Cross-references: GB:AE000165; GB:U00096; NID:g1786808; PIDN:AAC73694.1; PID:g17868

A:Experimental source: strain K-12, substrain MG1655

C:Comment: This enzyme catalyzes the isomerization of chorismate to isochorismate, an

of the entc gene is regulated by iron and possibly by the product of the enta gene, 2

C:Genetics:

A:Gene: entC; fepP; menP

A:Map position: 14 min

C:Function:

A:Description: EC 5.4.99.6 [validated; MUID:89123153]

A:Pathway: enterobactin biosynthesis; menaquinone biosynthesis

C:Superfamily: Isochorismate synthase

C:Keywords: enterobactin biosynthesis; intramolecular transferase; isomerase

Query Match 17.1%; Score 344; DB 1; Length 391;

Best Local Similarity 28.6%; Pred. No. 8.3e-19;

Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;

Db 19 YAALETTFGAYATP-----DRLRFGIGATATKTAQALOGAVF-GAOSFDEQEP 71

QY 72 QSELMAQFWVPEPMVITADKITFGSDIYSDFTTMAQVPRKQPNVTTSHTVDEWIT 131

Db 86 SS-----LYTPE-----SW--QSFSSRQEKQASRRFTR----- 111

QY 132 ERTENLIDTLAI--DQTLAKVVFGRQTLQIS--DTRLAQIIRALAEQANTYHVTKR- 186

Db 112 SQSLNVERQALPEQTTFEQMVARAAALITATPOYDKVLSRLIDITTDAAIDSGVLERL 171

QY 187 -----HDEL-----FISATPERLVAMSGOIATAAVAGTSRGTGADDIALGEA 231

Db 172 IAQNPVSYNFHVPLADGVLIGASPELLLRKRDGRFSSIPLAGSARQPEVDLREAGNR 231

QY 232 LMSQKRRIHQYVVAITRLQDVTSLSKVPAMSLKKNQVQHLVPTIGDIAHLVS 291

Db 233 LMSQKRRIHQYVVAITRLQDVTSLSKVPAMSLKKNQVQHLVPTIGDIAHLVS 291

QY 292 TAIYDRILHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 351

Db 292 LTIACLIHPPLAGVPREAALYYIATHEKTPRGRLAGPGYGTADNSGSEVY 351

QY 352 NOTORRATLFPAGAGIVADSDAQOEYEETGLKFEP 386

Db 352 RENQVR--LFGAGIVPASSPLGEMRETGVKLSM 384

RESULT 6

C69615 Isochorismate synthase dhbc - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: CG9615  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertelle, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.; Chen, C.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. *Nature* 390, 249-256, 1997  
A:Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallere, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Iech, J.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Koetter, P.; Lazarevic, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Authors: Lauber, J.; Lazarevic, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Pottebell, Rieger, M.; Rivolta, S.; Kocho, E.; Kocho, E.; Kocho, B.; Rose, M.; Sadale, J.; Sato, T.; Scanlon, A:Authors: Schliebl, C.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serotini, akuchil, M.; Tamakoshi, A.; Tanaka, T.; Terpilata, P.; Tsugonoi, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipal, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:96044033  
A:Accession: CG9615  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residue: 1-398 <RUN>  
A:Cross-references: GB:29120; GB:RL009126; NID:g2635613; PIDN:CAB15189.1; PID:g2635696  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: dhdc  
C:Superfamily: Isochorismate synthase

Query Match	16.4%	Score	328.5	DB 1	Length	398
Best Local Similarly	28.5%	Pred. No.	1.3e-17			
Matches 111; Conservative	64;	Mismatches	163;	Indels	51;	Gaps 12

[illegible]

RESULT 7  
A40365  
siderophore biosynthetic protein amoA - Aeromonas hydrophila  
C:Species: Aeromonas hydrophila  
C:Date: 10-Sep-1999 #sequenceRevision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A40365  
R:Barghouti, S.; Payne, S.M.; Arceneaux, J.E.L.; Byers, B.R.  
J. Bacteriol. 173, 5121-5128, 1991  
A:Title: Cloning, mutagenesis, and nucleotide sequence of a siderophore biosynthetic gene  
A:Reference number: A40365, MUID:91317731  
A:Accession: A40365  
A:Status: Preliminary

A, Molecule type: DNA  
A, Residues: 1-396 <BAR>  
A, Cross-references: GB:M63339  
A, Note: the authors translated the codon GAG for residue 393 as Gly  
C, Superfamily: isochoirismate synthase

Query Match	16.3%	Score	326.5	DB 1	Length	396			
Best Local Similarity	35.2%	Pred	No. 1.8e-17						
Matches	87	Conservative	43	Mismatches	112	Indels	5	Gaps	4

[illegible]

QY	384	EPMRQLL	390
		:	
Db	382	GTMLKAL	388

RESULT 8  
G64997  
isochorismate synthase (EC 5.4.99.6) - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: G64997; S68696; S61203  
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.: Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: G64997  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-356 <BLAT>  
A:Cross-references: GB:AA000316; GB:U00096; NID:g1788594; PIDN:AACT5325.1; PID:g178866  
A:Experimental source: Strain K-12, substrain MC1655  
R:Mueller, R.; Dahm, C.; Schulte, G.; Leistner, E.  
FEBS Lett. 378, 131-134, 1996  
A:Title: An isochorismate hydroxymutase isogene in *Escherichia coli*.  
A:Reference number: S68696; MUID:96140724  
A:Accession: S68696  
A:Molecule type: DNA  
A:Residues: 'MLTTRR', 30-93, 'E', 95-356 <MUE>  
A:Cross-references: EMBL:Z50849; NID:g1050906; PID:g1050907  
A:Experimental source: strain MC4100  
C:Genetics:  
A:Gene: menP  
C:Keywords: intramolecular transferase; isomerase

```

Query Match          15.8%: Score 317.5: DB 2: Length 356:
Best Local Similarity 30.3%: Pred. No. 7.6e-17:
Matches 106: Conservative 51: Mismatches 162: Indels 31: Gaps 10:

QY 59 VFGAQSFDSEQYPOSELMAGFWFVEPVMTIADITKTEGSPVSDPT-----TWLA 109
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db  19 IWGLNAFPDSQ-----GNLLPRLEKRRCCGKATLRLTLESLSLDHDAIQKKEFLA 70
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

QY 110 QFVPKQP-----NTVTTSHTVDEVDWTEKTEENLIDTLAIDQTLAKVVFGRQOTQLQSLDT 163
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

Db 71 TLVSTKPLRPGHLHTTTRRQHMPODKTGMTOJIELATKRIA-EGELDKVYLVARLARDLHFASS 129

Qy 164 LRLAQIIRALAE-QANTYHVVLKKNDE-LTISATPELYAMSGQIATAVAAGTSRRGTD 221

Db 130 VNAAAMAAASRRILNICYHFHFMARDGENAEFGSPERLMMRRDKALTELTALGTAVANNPD 189

Qy 222 GADDIALGELLASOKRIEHOYVVASITTRLDQVTSLLKVPMPSLKKKVOYHLYTP1 281

Db 190 DKAOQUGELMIMADKKQREMLVVEDICORLADTQTLV-LPPOYLIRKKQYHLLRCT 248

Qy 282 TGDIAAHLSTVTAIVDLRHLPTPALGSGVREALALYIATHEKTPRGFLFAGPIGYFTADNSGE 341

Db 249 WTSINKRADVDICL-HQLOPFAAIVAGLPRDLARQFIARHPPETREWAGASAGYLSLQGS-E 306

Qy 342 FVVGIRSMYVNOTORRATLRFAGAIYADSDAOEYETGKLFERPMOLK 391

Db 307 FCVSRIRSKISGNNVR-LVNAQAIIVGSDPEDEWQSIDKKAGLRLTL 354

```

RESULT      9
JC5323
anthranilate synthase (EC 4.1.3.27) component I [validated] - Sulfolobus solfataricus
N:Alternate names: anthranilate synthase alpha chain
C:Species: Sulfolobus solfataricus
C:Date: 15-May-1997 #sequence_revision 18-Jul-1997 #text_change 17-Mar-2000
C:Accession: JC5323; A40635
R:Tutino, M.L.; Tosco, A.; Martino, G.; Sanna, G.
Biochem. Biophys. Res. Commun. 230, 306-310, 1997
A:Title: Expression of Sulfolobus solfataricus trpE and trpG genes in E.coli.
A:Reference number: JC5323; MUID:97168965
A:Accession: JC5323
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-421 <?GT?>
R:Tutino, M.L.; Searno, G.; Marino, G.; Sanna, G.; Cubellis, M.V.
J. Bacteriol. 175, 299-302, 1993
A:Title: Tryptophan biosynthesis genes trpEC in the thermacidophilic archaeobacterium
A:Reference number: A40635; MUID:93106970
A:Contents: Mt-4
A:Accession: A40635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <?TU?>
A:Cross-references: GB:M98048; NID:g152936; PIDN:AAA73379.1; PID:g152937
A:Note: Sequence extracted from NCBI backbone (NCBIN:121234, NCBI:P.121235)
C:Genetics:
A:Gene: trpE
C:Complex: heterotetramer; two component I chains (PIR:JC5323), two component II chains
C:Function: <AMT?>
A:Description: EC 4.1.3.27 [validated; MUID:97168965]
A:Pathway: tryptophan biosynthesis
A:Note: first step
C:Function: <COM1>
A:Description: EC 4.1.3.27 [validated; MUID:99380543]
A:Note: magnesium required
C:Superfamily: anthranilate synthase component I
C:Keywords: carbon-carbon lyase; oxo-acid-lyase; tryptophan biosynthesis

Query Match      14.48;      Score 289; DB 2;      length 421;
Best Local Similarity 24.98;      Pred. No. 1.5e-14;
Matches 95; Conservative 72; Mismatches 163; Indels 52; Gaps 11;

OY      36 DDTLRFGIGATATKATQ--ALOGAVFGAQSPP-----EGEYQSELMAG 78
      || : : : : : : : : : : : : : : : : : : : : : : : : :
DB      56 DDPAVILNGYKDKLADIPGLFKGMIGYSDYAVRFEKIRDKLRAAEDWPAE--- 111
      || : : : : : : : : : : : : : : : : : : : : : : : : :

OY      79 FWEVP-----EVMVTIADKITEGSDTVSDFTTMAQFVYKQPNVTYTSHTDE-V 128
      || : : : : : : : : : : : : : : : : : : : : : : : : :
DB      112 -FFPDNIIITVDHNGKYYVADLSVGG--CGDIGEFKVSFYDESINKKSYERIVSESL 168
      || : : : : : : : : : : : : : : : : : : : : : : : : :

OY      129 DWIEETWMLITLAIIDTLAKVVFGR-QQTLQLSDTLRLAQIIPALAEQANTYHVLKRR 187
      || : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      169 EKIR-----SGEIQVLSRFYRIFFSDPLRIYNNLRIT--NSPFWELIKED 215

QY      188 DELFISATPERLIVAMSGQIATAAVAGTSRRGTGDADDIALGEALLASQNRRIEHQVYA 247
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      216 EKYLLGSSPELLFRQODINIVETPYIACIRPGAQOEEEDLKLELEIMSEKDKAEHLMLVD 275
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      248 SITTFLODQVTF--SLKIVAMPSLLKNNQVQCHLYPIPIQDIAHLSTVAIYDRLHPPALG 305
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      276 LARNDLGVQVPGVYKVPBELMYEKYSYHQAIVSKYIGTLKKKYNALNVLSTATFPAGTYS 335
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      306 GVPREALATYATIEHEKTPRFLAPDIGEFTYTDNDSGEVNGVIRSMYVNOTORATLEFAGG 365
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      336 GPRKPMANNITIELEEYRGPRGYPACAVGISADNGAAFEALINRTAFLEKELLR--IHAQNG 393
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      366 IYADSDAQOEYEETGKLEFPKR 387
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      394 IYDSDNPSEYFEETEHKLKALK 415
        :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT      10
A42301
anthranilate synthase (EC 4.1.3.27) alpha chain - Haloferax volcanii
C:Species: Haloferax volcanii
C:Date: 10-Jul-1992 #sequence_revision 18-Sep-1992 #text_change 22-Jun-1999
C:Accession: A42301
R:Lam, W.L.; Logan, S.M.; Doolittle, W.F.
J. Bacteriol. 174, 1694-1697, 1992
A:Title: Genes for tryptophan biosynthesis in the halophilic archaeobacterium Haloferax
A:Reference number: A42301; MUID:92165748
A:Accession: A42301
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <LB>
A:Cross-references: GB:M8738; NID:g149036; PID:AA73177.1; PID:g149039
A>Note: the authors failed to give the translation for ACC in residue 239 as shown in
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match      13.9%; Score 278.5; DB 2; Length 523;
Best Local Similarity 28.9%; Pred. No. 1,36-13;
Matches 79; Conservative 44; Mismatches 115; Indels 35; Gaps 5;

QY 120 TTSHTVDEVDW--IERTENL--IDTLAIDQTLAKVFGROQTLDSTLRLLAQIIRAL 173
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 256 FKEHVRDGDIQGVISFTRRLRGQVDPVGLVASREV----- 292

QY 174 AEQANTFIVVLKRHDELFSATPERLVAMSGQIATAVAAGTSRGRDGDADIALGELL 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 293 --NPSPYFELLRHGRRRVVGASPELVSVRGRDVAVNPVIACTCORGSGPVEDRLAELL 350

QY 234 ASQKRRIEQQVVASITTRLQDVT--SLAKPAMPSPILKNNQVOHLKPTITGDIAHLAV 291
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 351 ADAKERAHEHTMLVDLGRNDVRRVSTPGSVKREDENSIITKSHVOHISTVSGTLADADA 410

QY 292 TAIIVRLHPTALGVPPEALALYYATHEKIPRGLFAGPIGTFYADNSGEFVVGIRSMYV 351
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 411 FQATATPAGLITGAPKVRAMEIIDDLEAPRGVYGGGVGYISWTGDADVAIYIRATYV 470

QY 352 NO--TQRRATLFGAGIVADSDAQOEYEETGLK 382
      : : : : | | | | | | | | | | | | | | | | | | | | | |
DB 471 DSGADDAITVRAGAGIVADSDPTAEYEETQK 503

RESULT      11
E64059
probable isochorismate synthase (EC 5.4.99.6) - Haemophilus influenzae (strain Rd KW2
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 16-Jul-1999
C:Accession: E64059
R:Flitschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
J. Genome, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
J. D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

```





[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 02:18:04 ; Search time 47.14 Seconds  
(without alignments)  
269,057 Million cell updates/sec

Title: US-08-774-104A-2

Perfect score: 1 MMTYHETRALAQSDLOQLYA.....ETGLKFEPMRLQLDYNHVE 397  
Sequence: 2009

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496.5	24.7	471	1	MENF_BACSU
2	453	22.5	476	1	PCNA_PSEAE
3	344	17.1	391	1	ENHC_ECOLI
4	328.5	16.4	398	1	DHBC_BACSU
5	326.5	16.3	396	1	AMOA_AERHY
6	317.5	15.8	431	1	MENF_ECOLI
7	308.5	15.4	395	1	VIBC_VIBCH
8	289	14.4	421	1	TRPE_SUISO
9	278.5	13.9	523	1	TRPE_HALVO
10	278	13.8	430	1	MENF_HAFIN
11	272.5	13.6	456	1	TRPE_LACLA
12	257	12.8	516	1	TRPE_MYCTU
13	253	12.6	511	1	TRPE_STRKO
14	252.5	12.6	433	1	TRPE_PYRKO
15	249.5	12.4	470	1	PABB_BACSU
16	246.5	12.3	494	1	TRPE_AODAE
17	246	12.2	464	1	TRPE_METH
18	245	12.2	461	1	TRPE_THEMA
19	244.5	12.2	515	1	TRPE_BACSU
20	244.5	12.2	513	1	TRPE_BACPU
21	244	12.1	462	1	TRPE_METPM
22	236	11.7	529	1	TRPE_MYCLE
23	234.5	11.7	478	1	TRPE_AERPE
24	232	11.5	434	1	TRPE_METJA
25	226	11.2	531	1	TRPE_ARTGO
26	220	11.0	497	1	TRPE_PACICA
27	219	10.9	508	1	TRPE_BACST
28	219	10.9	508	1	TRPE_SYNK3
29	217.5	10.8	454	1	PABB_SALTY
30	217	10.8	454	1	TRPE_CLOTM
31	217	10.8	454	1	TRPE_BACCA
32	215	10.7	485	1	TRF2_SYNS3
33	214	10.7	500	1	TRPE_RHOSH

34	213.5	10.6	411	1	TRPE_ARCFU	028669 archaeoglob
35	213.5	10.6	506	1	TRPE_YEAST	P00899 saccharomyc
36	211.5	10.5	491	1	TRPE_NEIMB	P56995 neisseria m
37	210.5	10.5	491	1	TRPE_NEIMA	Q9xaz0 neisseria m
38	210.5	10.5	491	1	TRPE_NEIMC	Q9s358 neisseria m
39	210.5	10.5	595	1	TRPE_ARATH	P2068 arabisdopsis
40	209.5	10.4	453	1	PABB_ECOLI	P05041 escherichia
41	209	10.4	462	1	TRPE_THETH	P05378 thermus aqu
42	208.5	10.4	491	1	TRPE_NEIGO	Q9w400 neisseria g
43	208.5	10.4	492	1	TRPE_PSEAE	P20580 pseudomonas
44	201.5	10.0	621	1	TRPX_ARATH	P32069 arabidopsis
45	200.5	10.0	451	1	PABB_KLEAE	P12679 klebsiella

## ALIGNMENTS

```

RESULT 1
ID      MENF_BACSU      STANDARD;      PRT:      471 AA.
AC      P23973; P23972;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DE      15-DEC-1998 (Rel. 37, Last annotation update)
DE      MENAQUINONE-SPECIFIC ISOCORISMATE SYNTHASE (EC 5.4.99.6).
GN      MENF OR ICSM.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/staphylococcus group; Bacillus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-168 / RB1.
RX      MEDLINE; 96144257.
RA      Rowland B., Hill K., Miller P., Driscoll J.R., Taber H.W.;
RT      "Structural organization of a Bacillus subtilis operon encoding
RT      menaquinone biosynthetic enzymes.";
RL      Gene 167:105-109(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 98048467.
RA      Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT      "Sequencing and functional annotation of the Bacillus subtilis genes
RT      in the 200 kb rnb-dnaB region.";
RL      Microbiology 143:3431-3441(1997).
RN      [3]
RP      PRELIMINARY SEQUENCE FROM N.A.
RC      STRAIN-168 / RB1;
RX      MEDLINE; 92332443.
RA      Driscoll J.R., Taber H.W.;
RT      "Sequence organization and regulation of the Bacillus subtilis menB
RT      operon.";
RL      J. Bacteriol. 174:5063-5071(1992).
RN      [4]
RP      SEQUENCE OF 1-32 FROM N.A.
RX      MEDLINE; 88227858.
RA      Miller P., Mueller J., Hill K., Taber H.W.;
RT      "Transcriptional regulation of a promoter in the men gene cluster of
RT      Bacillus subtilis.";
RL      J. Bacteriol. 170:2742-2748(1988).
CC      -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCORISMATE.
CC      -1- PATHWAY: MENAQUINONE BIOSYNTHESIS.
CC      -1- SIMILARITY: STRONG, TO OTHER ISOCORISMATE SYNTHASES; WEAK, TO
CC      TRPE AND PABB.
CC      -1- CAUTION: USED TO INCLUDE WHAT WAS CALLED 'MENR'.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

DR EMBL; M74538; AAC37013.1; -.  
 DR EMBL; AF008220; AAC00223.1; -.  
 DR EMBL; M74521; AAA50396.1; ALT\_SEQ.  
 DR EMBL; M74521; AAA50397.1; ALT\_SEQ.  
 DR EMBL; M21320; AAA22594.1; -.  
 DR EMBL; Z99119; CAB15061.1; -.  
 DR PIR; S27508; S27508.  
 DR PIR; S27507; S27507.  
 DR SUBTILIST; BG10682; MENF.  
 DR INTERPRO; IPR000350; -.  
 DR PFAM; PF00425; chorismate\_bind; 1.  
 DR Menaglutone biosynthesis; Isomerase.  
 FW CONFLICT 12 E -> K (IN REF. 4).  
 KW SEQUENCE 471 AA; 52811 MW; 94C047C549FC4CF9 CRC64;

Query Match 24.7%; Score 496.5; DB 1; Length 471;  
 Best Local Similarity 30.2%; Pred. No. 1.9e-29;  
 Matches 127; Conservative 77; Mismatches 161; Indels 55; Gaps 10;

19 YAALETTERGAFATPADDTLRFGIGALITAKTAQ----- 53  
 46 YGAKYITGNRFWSDESELTIVGLKEAVFOTNKNRKYREVEEQWERFKTAHIE 105  
 54 -----ALQAVFGAQSFEDEYPOSE---LMAGFWPEVYVNTIADK-----IT 95  
 106 EEKLGHSAGVPLFGGFSFDPCEERGSQMDHFSSEGDFEVPALMTLTATGPFITVNRWS 165  
 96 FGSD---IVSDPTTLAOF-VP-KQPN--YVTTSHVDEVMVIERTEMLIDTLAIDQTL 147  
 166 GGEDEAVALEGLKAPAEWMPDFKQEDQAVTAAAEELDKDWLKRIEATSOIK-EKQY 224  
 148 AKVVGROQTLOSLDRLAOLIRALA-EOANTYHVLRKHDELFIATPERLVAMSGQ 206  
 225 DKVTLARELLTFDPRIQIEPVKLTLLDDQTSYFAIEQEGKTYGASPERLIKDDGT 284  
 207 TATAVAGTSRRGTGADIDALGEALLASQKNRIEHQYVVASITRLODVTSLKVPAMP 266  
 285 VMSCLAGSIRKGVNEEDDRIGLELLENDEKMLEHDIVGMHNAFVSSCSEVEKPDGP 344  
 267 SLKKKQVOHLYPTITGDIAHLSTAIYDRLHPTRPALGQPREALYIATHEKTPRL 326  
 345 VLYKTSVOHLETPVIGQRESASIFDLIEKLPALGAGSPQEKAVDVIETIEPMSRGM 404  
 327 FAGPIGYFADSGFEVVGIRSMVYNOTORATLFAAGIVADSDAQOEFEETGLKEPM 386  
 405 YAPIGWIDSQNGEFNAIRSGLIEGTAR--LRAGCIVDESEPTISEYETQIKLKM 462

RESULT 2  
 PCHA\_PSEAE STANDARD: PRT; 476 AA.  
 AC Q51508; 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).  
 GN PCHA.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE; 96086939.  
 RA Setino L., Reimann C., Baur H., Beyeler M., Visca P., Haas D.;  
 RT "Structural genes for salicylate biosynthesis from chorismate in  
 RT Pseudomonas aeruginosa."  
 RL Mol. Gen. Genet. 249:217-228(1995).  
 CC -I- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE  
 CC (PROBABLE).  
 CC -I- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
 CC -I- PATHWAY: SALICYLATE BIOSYNTHESIS.

CC -I- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 CC TYPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X82644; CAAS7969.1; -.  
 DR INTERPRO; IPR000350; -.  
 DR PFAM; PF00425; chorismate\_bind; 1.  
 KW Isomerase.  
 SO SEQUENCE 476 AA; 52071 MW; D0F6AFD9C9CFC9F5 CRC64;

Query Match 22.5%; Score 453; DB 1; Length 476;  
 Best Local Similarity 42.4%; Pred. No. 3.1e-26;  
 Matches 103; Conservative 38; Mismatches 98; Indels 4; Gaps 3;

149 KYVFGQQTLOSLDRLAOLIRAL-AEOANTYHVLRKHDELFIATPERLVAMSGQI 207  
 221 KYVLTARTQARPLGD-IEPQVIEHLRLQHADQLFACRGRGNCFLGASPERIVRIRAGA 279  
 208 ATAAVAGTSRRGTGADIDALGEALLASQKNRIEHQYVVASITRLODVTSLKVPAMP 267  
 280 LTHALGTLTARGDQADRLGALDLSAKDRHEHLYVEALITALEPSEVLEITDAG 339  
 268 LKKKQVOHLYPTITGDIAHLSTAIYDRLHPTRPALGQPREALYIATHEKTPRL 327  
 340 LKRLAVQHLNPTIRARLADAGILRLQALHPTPAVGVPRAALDYIRQHGMDRGW 399  
 328 AGPIGFTFDNNGSEFVVGIRSMVYNOTORATLFAAGIVADSDAQOEFEETGLKEPM 387  
 400 AAPLGWLDEGNDPFLVALRSALL--TPGRGYLFACGIGVSEPAHEVRETCLKLSMR 457  
 388 QLL 390  
 458 EAL 460

RESULT 3  
 ENTC\_ECOLI STANDARD: PRT; 391 AA.  
 AC P10377; 01-MAR-1989 (Rel. 10, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ISOCHORISMATE SYNTHASE ENTC (EC 5.4.99.6).  
 GN ENTC.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE; 90236256.  
 RA Elkins M.F., Earhart C.F.;  
 RT "Opacity factor from group A streptococci is an apoproteolase."  
 RL FEMS Microbiol. Lett. 56:35-40(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE; 89123153.  
 RA Ozenberger B.A., Brickman T.J., McIntosh M.A.;  
 RT "Nucleotide sequence of Escherichia coli isochorismate synthetase  
 RT gene entc and evolutionary relationship of isochorismate synthetase  
 RT and other chorismate-utilizing enzymes."  
 RL J. Bacteriol. 171:775-783(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE: 90230305.
RA Brickman T.J., Ozenberger B.A., McIntosh M.A.;
RT "Regulation of divergent transcription from the iron-responsive fepr-
RT entc promoter-operator regions in Escherichia coli.";
RL J. Mol. Biol. 212:669-682(1990).
RN [5]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE: 90241936.
RA Liu J., Quinn N., Berchtold G.A., Walsh C.T.;
RT "Overexpression, purification, and characterization of isochorismate
RT synthase (entC), the first enzyme involved in the biosynthesis of
RT enterobactin from chorismate.";
RL Biochemistry 29:1417-1425(1990).
CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.
CC -1- PATHWAY: ENTEROBACTIN BIOSYNTHESIS. ENTEROBACTIN IS AN IRON-
CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE
CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.
CC -1- SUBUNIT: MONOMER.
CC -1- INDUCTION: EXPRESSED UNDER CONDITIONS OF IRON STARVATION.
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TRPE AND PABB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24142; AAA16100.1; -
DR EMBL: M36700; AAA18491.1; -
DR EMBL: AE000165; AAC73694.1; -
DR EMBL: X53274; CAA37371.1; -
DR PIR: J70497; SEYCIK.
DR PIR: S04323; S04323.
DR ECOGENE: EG10261; ENTC.
DR INTERPRO: IPR000350; -
DR PFAM: PF00425; chorismate_bind; 1.
DR Enterobactin biosynthesis; Iron transport; Isomerase.
DR KW CONFLICT 305 306 SG -> TA (IN REF. 1).
SQ SEQUENCE 391 AA; 42931 MW; 62862569DFC41AC4 CRC64;

```

Query Match 17.1%; Score 344; DB 1; Length 391;  
 Best Local Similarity 28.6%; Pred. No. 2.6e-18;  
 Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12;

```

QY 232 LIAOKRIEHOVVASITTRLDVNTSLKVPAMSLKKNQVOHLYPTIGDIAHLNV 291
DB 232 LIASEKDRHEHELTQMKKEYLRERSSSELHPSSPOLITPTLWHLATPEGKANSOENA 291
QY 292 TAIYDRIDHPTPALGAGVPREALYIATHEKTPRGLFAGPIGYPADSGEFVNGIRMYV 351
DB 292 LTLACLHPTPALSGFPHQAATQVIALEPPDRRLFGGIYGMCDSENGENVYIRCAKL 351
QY 352 NOTORATLFGAGIVADSDPAQOEYEETGLKFEPM 386
DB 352 RENQVR--LEFAGAGIVPASSPLGEMRETFVGLSTM 384
RESULT 4
DHC-BACSU STANDARD; PRT; 398 AA.
AC P45744;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ISOCHORISMATE SYNTHASE DHC (EC 5.4.99.6).
GN DHC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-168 / MARBURG;
RX MEDLINE: 96146538.
RA Rowland B.M., Taber H.W.;
RT "Duplicate isochorismate synthase genes of Bacillus subtilis:
RT regulation and involvement in the biosyntheses of menaquinone and
RT 2,3-dihydroxybenzoate.";
RL J. Bacteriol. 178:854-861(1996).
CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.
CC -1- PATHWAY: 2,3-DIHYDROXYBENZOATE BIOSYNTHESIS.
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TRPE AND PABB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U26444; AAC44631.1; -
DR EMBL: Z99120; CAB15189.1; -
DR SUBTILIST; BG11242; DHC.
DR INTERPRO: IPR000350; -
DR PFAM: PF00425; chorismate_bind; 1.
DR Isomerase.
SQ SEQUENCE 398 AA; 43446 MW; 3311FB3517C79BC CRC64;

```

Query Match 16.4%; Score 328.5; DB 1; Length 398;  
 Best Local Similarity 28.5%; Pred. No. 3.7e-17;  
 Matches 111; Conservative 64; Mismatches 163; Indels 51; Gaps 12;

```

QY 28 GAYF-ATPADDTLRFGGAIA-----TAKRQAOLQAGVFGAQSDEDEYQSE 74
DB 22 GAFFLASPHRYLLAKGICEIVPEADGONQMETLISGRIAEALROAKOSGQS----- 71
QY 75 LMAGFVPEVMVNTIADKITFGSDTVSDFTWYLA--QF--VPROPTNTVTSH--VTDE 127
DB 72 -----RPLVGVAVPPDQYKAARLVVPEVYRWGSPLOTFHEEKEQAGHYHKKPVP 127
QY 128 VDWIERTENLIDTLAIDQTLAKVFGSQDTLQSLDTRLAQIIRALAEQAN--TYHVL 184
DB 125 EDYKNGVEQGLARIA-DGTLISKIVLSRSLHLSPEPIQTDLRLHQAQHNHSHGYTFADV 183

```

[illegible]

```

Db      264 SPSLMSTDTLWHLGTPPIAGRLNGCEASVSIACOLHPTPLCGYPIELARQFIREDEOPR 323
Qy      324 RGLFAGPFGYTPAANSGEFFVVGISMVYNOTRRATFIFAGAGIVADSDAOQEYEETGLAF 383
Db      324 RALFSGIYGVWCDSDQNGEMAVVINGVLDGHQ--VELFAGAGIVASGDPWMSGPRRGTRL 381
Qy      384 EPMROL 390
Db      382 GTMLKAL 388

RESULT 6
MENF_ECOLI STANDARD; PRT; 431 AA.
ID MENF_ECOLI
AC P38051; Q47009; Q47704; P76479; P78297;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DI MENAQUINONE-SPECIFIC ISOCHORISMATE SYNTHASE (EC 5.4.99.6).
GN MENF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MC4100;
RX MEDLINE; 96140724.
RA Mueller R., Dahm C., Schulte G., Leisner E.;
RT "An isochorismate hydroxymutase isogene in Escherichia coli.";
RN FEBS Lett. 378:131-134(1996).
RN [2]
RP REVISIONS.
RA Mueller R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / PL2024;
RX MEDLINE; 96291661.
RA Daruwala R., Kwon O., Meganathan R., Hudspeth M.E.S.;
RT "A new isochorismate synthase specifically involved in menaquinone
(vitamin K2) biosynthesis encoded by the menf gene.";
RN FEBS Microbiol. Lett. 140:155-163(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 1-216 FROM N.A.
RC STRAIN-K12;
RA Hulsman G.W.;
RN Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 216-431 FROM N.A.
RC MEDLINE; 89327151.
RA Pop J.L.;
RT "Sequence and overexpression of the menf gene from Escherichia coli."

```

RL J. Bacteriol. 171:4349-4354(1989).  
RN [8]  
RP IDENTIFICATION  
RX MEDLINE: 95075659.  
RA Borodovsky M., Rudd K.E., Koonin E.V.;  
RT "Intrinsic and extrinsic approaches for detecting genes in a  
bacterial genome.";  
RL Nucleic Acids Res. 22:4756-4767(1994).  
CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
CC -1- PATHWAY: MENADIOLONE BIOSYNTHESIS.  
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
CC TRPE AND PABP.  
CC -1- CAUTION: REF. 7 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
CC FRAMESHIFTS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z50849; CAA90702.1; -  
DR EMBL: U54790; AAC44303.1; -  
DR EMBL: AE000316; AAC75325.1; ALT\_INIT.  
DR EMBL: D90857; CAB22015.1; -  
DR EMBL: D90858; CAB22025.1; -  
DR EMBL: U58768; AAB02729.1; -  
DR EMBL: M21787; -; NOT\_ANNOTATED\_CDS.  
DR ECOGENE: EG12362; MENF.  
DR INTERPRO: IPR000350; -  
DR PFM: PF00425; chorismate\_bind; 1.  
DR Menadiolone biosynthesis; Isomerase.  
KW CONFLICT 310 312 VLR -> YC (IN REF. 3).  
SQ SEQUENCE 431 AA; 48764 MW; D081724F3329FE47 CRC64;  
  
Query Match 15.8%; Score 317.5; DB 1; Length 431;  
Best Local Similarity 30.3%; Pred. No. 2.7e-16;  
Matches 106; Conservative 51; Mismatches 162; Indels 31; Gaps 10;  
  
QY 59 VFQASFEDEQEYQSELMAGFWFVEVAVTIAADKIFGSDYSDPT-----TWLA 109  
DB 94 IMLGNMFDSQ-----GNLLPRLKMRGCGKATRLTLTSESSLDHDAIDAKFEITA 145  
QY 110 QFVPRKP-----NTVYTSHTVDEVMIEKTEMLDITLADITLAKVVFGRQOTLQSLDT 163  
DB 146 TLVSIRKPLRLHLLTTRQGWMPDKTGWTLDELATKTLA-EGELDKVVLARATDLHFRASP 204  
QY 164 LRLAQITRALAE-QANTYHVVLKRHDE-LFISATPERVLAMSGGQITAAVAGTSRSGTD 221  
DB 205 VNAAMAAARRLNLCNYHYMAFDGEMAFGLSGSPERLMRRDKALRTEALAGTVANNPD 264  
QY 222 GADDIALLGEALLSOKKRIRIEHGYVVASITTRLDVTSLSKVPAMPSSLLKKOVOLHTPI 281  
DB 265 DKQAQOLGEMLMDKKQREMLVEDICQLQADTOTLTV-LPPQVLRLLKVOHLRRCI 323  
QY 282 TGDIAHLSTAIIVRLHPTPALGVPREALAYTATHEKTRPGLGPIGYFTADNSGE 341  
DB 324 WTSLSKNADVDYCL-HQOLPFAAVALGRDLARQFIARHEPFRTRMYVAGSYLSLOS-E 381  
QY 342 FVVGISMTYVNOTORRATLEFAGAGIVADSPAQOEYEETGKFERPQRLK 391  
DB 382 FCVLSRSKISGNVVR--LYAGAGIVGSDPEQEMOEIDNKAGLRLTLQ 429  
  
RESULT 7  
VIBC\_VIBCH  
ID VIBC\_VIBCH STANDARD; PRT; 395 AA.  
AC 007898;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE VIBRIOBACTIN-SPECIFIC ISOCHORISMATE SYNTHASE (EC 5.4.99.6).  
GN VIBC.  
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN: EL TOR / L0015;  
RX MEDLINE: 98037504.  
RA Wyckoff E.E., Stoeckner J.A., Reed K.E., Payne S.M.;  
RT "Cloning of a Vibrio cholerae vibriobactin gene cluster:  
RT identification of genes required for early steps in siderophore  
RT biosynthesis.";  
RL J. Bacteriol. 179:7055-7062(1997).  
CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.  
CC -1- PATHWAY: VIBRIOBACTIN BIOSYNTHESIS. VIBRIOBACTIN IS AN IRON-  
CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE  
CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.  
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
CC TRPE AND PABP.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U52150; AAC45925.1; -  
DR INTERPRO: IPR000350; -  
DR PFM: PF00425; chorismate\_bind; 1.  
DR Iron transport; Isomerase.  
KW SEQUENCE 395 AA; 43594 MW; 2FFAF87DD948C78B CRC64;  
  
Query Match 15.4%; Score 308.5; DB 1; Length 395;  
Best Local Similarity 28.5%; Pred. No. 1.1e-15;  
Matches 109; Conservative 57; Mismatches 158; Indels 59; Gaps 11;  
  
QY 30 YFATPADDTLRFGIG-----AIAATAKTA---QALOGAVFGAQSEFDEYQSELMAGFW 80  
DB 28 FFAASNSMLGCGVAHAFQQAIPPAELANAKKOLLQA-----KRDECNDP--LHGI- 78  
QY 81 FVPEVAVTIAADKIFGSDYSDPTWLAQFVPRQPMVTTSHTVDEVMIEKTEMLDIT 140  
DB 79 -----VPRPKPTPRMIRPRLTVSSPRLNRRPRLHRLQV-----AKLSS 119  
QY 141 LAIDQ-----TLAKVVFGRQOTLQSLDTLRLAQILR--ALAEQANTYHV 182  
DB 120 PSGEYKQGVSHLNMENHSGLSKVLRSRALEITEDEIALPTLLRSLLAIHNGYFAA 179  
QY 183 VLKRHDELFSATPERVLAMSGGQITAAVAGTSRSGTDGADDIALLGEALLSOKKRIR 242  
DB 180 SLDEQRKL-IGASPELLVAKRGNTLISNPLAGSRPSODAOENQRRASLNTAKDHEH 238  
QY 243 QYVVASITTRLDVTSLSKVPAMPSSLLKKOVOLHTPIITGDIA-AHLSTAIIVRLHPT 301  
DB 239 GLVVEEERIMSRCLNLTTPMVPVSVETETMLHLSTLLEQGVSDPEYCALQVAAADLHPT 298  
QY 302 PALGCVPREALAYTATHEKTRPGLFAGPIGYFTADNSGEFVVGIRSYVNOTORRATLE 361  
DB 299 PAVGCFPRESAYQAIREFLEEDRDGYFGWGVGCDARGNGEVVYIRCAEVGSHQMK--LF 356  
QY 362 AGAGIVADSPAQOEYEETGKFE 384  
DB 357 AGAGIVDESLPQSELETGAKMK 379  
  
RESULT 8  
TRPE\_SULSO  
ID TRPE\_SULSO STANDARD; PRT; 421 AA.

```

AC 006128;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 49255 / MT4;
RA MEDLINE: 93106970.
RT Tullino M.L., Scarano G., Marino G., Sanna G., Cubellis M.V.;
RT "Tryptophan biosynthesis genes trpEGC in the thermoacidophilic
RT archaeobacterium Sulfolobus solfataricus.";
RL J. Bacteriol. 175:299-302(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 49255 / MT4;
RA Tullino M.L., Cubellis M., Sanna G., Marino G.;
RT "The tryptophan operon in Sulfolobus solfataricus.";
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA MEDLINE: 99380543.
RT Knoechel T., Ivens A., Hester G., Gonzalez A., Bauerle R.,
RT Wilmanns M., Kirschner K., Jansoni J.N.;
RT "The crystal structure of anthranilate synthase from Sulfolobus
RT solfataricus: functional implications.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9479-9484(1999).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
CC PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II.
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M98048; AAA7379.1; -
DR EMBL: Z50014; CAA90311.1; -
DR PIR: A40635; A40635.
DR PDB: 1ODL; 18-AUG-99.
DR INTERPRO: IPR000350; -
DR PFM: PF00425; chorismate_bind.1.
DR PRINTS: PR00095; ANTSNTHASEI.
DR Tryptophan biosynthesis; Lyase; 3D-structure.
KW SEQUENCE 421 AA; 47736 MW; 0E5B9E776E298C61 CRC64;
SQ

```

```

DB 169 EYIR-----SGYIFQVVLRSFRYIFSGDPLRIYNNLRRI--NPSPYMFYLFKD 215
QY 188 DELFISATPERLIYAMGGQATATAVAGTSRGRDGDADIALGALLASCKNRHEHQYVA 247
DB 216 EYTLGSSPELLRVDNDIVETYPDIAGTRPRGADQEDDLLELNMSEKDKAHEMLVD 275
QY 248 SITTRQDVVT--SLKVPAMPSILKNQVOHLTPITGDIAMHLSTAYAIYDRHPTPALG 305
DB 276 LANLDGKVCVPGCTVAPPELMYVEKSHOHIYSYIGTLKKRYNNLNVLSATFPAGTYS 335
QY 306 GVPREALYIATHEKTPNGLEFAGPIGYFTADNSGEFVNGIRSMYVNGQTRATLPAGAG 365
DB 336 GRKPPAMANNIEETLEEKKGPRYAGAVGISADGNAEFAIARPAFLNKKELLR--IHAGAG 393
QY 366 IVADSDAQOEYEETGKLFPEMR 387
DB 394 IVYDSNPESEYFETHEKALK 415

RESULT 9
TREE HALVO
ID TRPE HALVO STANDARD; PRT; 523 AA.
AC P33975;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE.
OS Halobacterium volcanii (Haloflex volcanii).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloflex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WED11;
RA MEDLINE: 92165748.
RT Lam W.L., Logan S.M., Doolittle W.F.;
RT "Genes for tryptophan biosynthesis in the halophilic archaeobacterium
RT Haloflex volcanii: the trpFEG cluster.";
RL J. Bacteriol. 174:1694-1697(1992).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
CC PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: TETRAMER OF TWO COMPONENTS I AND TWO COMPONENTS II (BY
CC SIMILARITY).
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M83788; AAA73177.1; -
DR PIR: A42301; A42301.
DR INTERPRO: IPR000350; -
DR PFM: PF00425; chorismate_bind.1.
DR PRINTS: PR00095; ANTSNTHASEI.
DR Tryptophan biosynthesis; Lyase.
KW SEQUENCE 523 AA; 55920 MW; 9765667997636A6C CRC64;
SQ

```

```

Query Match 14.4%; Score 289; DB 1; Length 421;
Best Local Similarity 24.9%; Pred. No. 3.3e-14;
Matches 95; Conservative 72; Mismatches 163; Indels 52; Gaps 11;

```

```

Query Match 13.9%; Score 278.5; DB 1; Length 523;
Best Local Similarity 28.9%; Pred. No. 2.6e-13;
Matches 79; Conservative 44; Mismatches 115; Indels 35; Gaps 5;

```



QY 174 AEOANTHVAVIKRDELFIATPERLVAMSGOATATAVAAGTSRGRDGDADIALGEALL 233  
 DB 293 --NPSPEMFLRHDGRVYVAGSPETLVSVKDRVYVNPITGTCORGSPVEDRRLABELL 350  
 QY 234 ASQKNRIEHOYVAVSITTRLODYTT--SLKVPAMPSSLKKNVOHLYPTITGDIAHLSTV 291  
 DB 351 ADKEREHEMTMLVDLGNDRVSTPGSVREDEPMISIKTSVQHISSTVSGTLDADADA 410  
 QY 292 TAYVDRHPPALGCVREALTYIATHEKTPRGLFAGPIGFTADNSGGEVVGIRSMYV 351  
 DB 411 FDATRATPAGTLLGAKFVAMEIIDLLEAPRGVGGVGYVSWTGDADVAIVIRATV 470  
 QY 352 NO--TORRATLFPAGAGIVADSDAQOEEYENGK 382  
 DB 471 DSGADDAITVRAGAGIVADSDPAETEEYEOK 503  
 RESULT 10  
 MENF\_HAEIN  
 ID MENF\_HAEIN STANDARD; PRT; 430 AA.  
 AC P44613;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE MENAQUINONE-SPECIFIC ISOCHORISMATE SYNTHASE (EC 5.4.99.6).  
 GN MENF OR H10285.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20;  
 RX MEDLINE: 95350630.  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 influenzae Rd." (1995).  
 RL Science 269:496-512(1995).  
 CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.  
 CC -1- PATHWAY: MENAQUINONE BIOSYNTHESIS.  
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO  
 TRPE AND PABB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U32714; AAC21947.1; -  
 DR TIGR: H10285; -  
 DR INTERPRO: IPR000350; -  
 DR PIRAM: PF000425; chorismate\_bind; 1.  
 KW Menadiquinone biosynthesis; isomerase.  
 SQ SEQUENCE 430 AA; 48466 MW; 3BA131F8B62A1B38 CRC64;

Query Match 13.8%; Score 278; DB 1; Length 430;  
 Best Local Similarity 27.8%; Pred. No. 2.2e-13;  
 Matches 101; Conservative 66; Mismatches 144; Indels 52; Gaps 17;

DB 68 ALAALGAV---QSFQSLNAQEFIEESGFLVVGGLQFGTQNOFVLPKMLVEQDNKGL- 122  
 QY 92 DKTF-----GSDIYSDFETW----LAOFVPPQPNVTTSHTYDEV--DWIERTEL 138  
 DB 123 --VSFEVKNQESANDTLAHLKTFENLTAALPQIPLHETLRANERTWCWVQA--LV 178  
 QY 139 DTLADTLAKVVGCGROOTLQSDTLRLAQITRLAEOAN--TYHV-LKRHDELFIAT 195  
 DB 179 EIKSGE--LTKIVLANETTFHLKQAINAYDFL-ASEKQNGCGHFLMAENSHSVFVST 235  
 QY 196 PERLVAMSGOATATAVAAGTSRRCTGDADIALGALLASQKNRIEHOYVAVSITTRLOD 255  
 DB 236 PERLFAREYNLLTEALAGTASVSESEFEPTQSANWMLLENDEKNLKENMLVEDISQNLK 295  
 QY 256 VTSLSKVPAMPSSLKKNVOHLYPTITGDIAHLSVAYIDRLAPTPALGCVPREALY 315  
 DB 296 QVESFVDSNV-ELKPLRVVQHLIRKIRANLFAHADVNIILKAHPHTAASGLPOQAKMI 354  
 QY 316 IATHEKTPRGLFAGPIGFTADNSGEFVYVGNVNOTORATLFPAGAGIVADSDAQOE 375  
 DB 355 LSEIETFDGWTAGTLGM-SDVCSEFCVAIRSAFIE--GHRINVFAGAGIVAGSQPLEE 411  
 QY 376 YEE 378  
 DB 412 WKE 414  
 RESULT 11  
 TRPE\_LACLA  
 ID TRPE\_LACLA STANDARD; PRT; 456 AA.  
 AC Q02001;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).  
 GN TRPE.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IL1403;  
 RX MEDLINE: 93015708.  
 RA Bardowski J., Ehrlich S.D., Chopin A.;  
 RT "Tryptophan biosynthesis genes in Lactococcus lactis subsp. lactis".  
 RL J. Bacteriol. 174:6563-6570(1992).  
 CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +  
 CC PYRUVATE + L-GLUTAMATE.  
 CC -1- SUBUNIT: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
 CC -1- SUBUNIT: Tetramer of two components I and two components II (by  
 CC similarity).  
 CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE  
 CC USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES  
 CC GLUTAMINE AMIDOTRANSFERASE ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M87483; AAA25223.1; -  
 DR PIR: S35124; S35124.  
 DR INTERPRO: IPR000350; -  
 DR PIRAM: PF000425; chorismate\_bind; 1.  
 DR PRINTS: PR00095; ANTSNTHASEI.  
 KW Tryptophan biosynthesis; lyase.  
 SQ SEQUENCE 456 AA; 51837 MW; 8CDEC9D949BD55C3 CRC64;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2000, 02:17:04 ; Search time 54.8 Seconds  
(without alignments)  
676.451 Million cell updates/sec

Title: US-08-774-104A-2  
Perfect score: 2009  
Sequence: 1 MWTYHETRALAQSDIQQLYA.....ETGLKEFPMQLLDKNHVE 397

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters:  297973
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1998.5	99.5	398	4	Q9UIA3	Q9UIA3 homo sapien
2	383	19.1	474	2	P74053	P74053 synechocyst
3	360	17.9	580	10	Q9ZPC0	Q9ZPC0 catharineti
4	354.5	17.6	372	2	O05851	O05851 mycobacteri
5	344	17.1	395	2	P77099	P77099 escherichia
6	307.5	15.3	503	10	O61522	O61522 arabidopsis
7	307.5	15.3	503	10	Q95788	Q95788 arabidopsis
8	304.5	15.2	391	2	P95475	P95475 pseudomonas
9	295.5	14.7	449	8	Q9TMO9	Q9TMO9 cyanidium c
10	251.5	12.5	532	1	O9V105	O9V105 pyrococcus
11	227	11.3	404	2	O9S202	O9S202 streptomyces
12	219	10.9	719	2	P72539	P72539 streptomyces
13	218.5	10.9	489	3	O94582	O94582 schizosacch
14	217.5	10.8	434	3	O9X918	O9X918 yersinia en
15	216.5	10.8	434	2	O9Z396	O9Z396 yersinia pe
16	216	10.8	434	2	O9Z6F1	O9Z6F1 yersinia pe
17	215.5	10.7	638	2	O9RXV9	O9RXV9 deinococcus
18	215	10.7	503	2	O9RQG2	O9RQG2 rhodobacter
19	214.5	10.7	473	2	Q9RTH3	Q9RTH3 deinococcus

20	210.5	10.5	364	3	P78905	P78905	schizosacch
21	202.5	10.1	522	2	Q9RQ44	Q9RQ44	buchnera ap
22	202.5	10.1	613	10	Q41156	Q41156	ruta grave
23	199.5	9.9	450	2	P71758	P71758	mycobacteri
24	199.5	9.9	608	10	Q41155	Q41155	ruta grave
25	195	9.7	616	10	Q81533	Q81533	nicotiana
26	189.5	9.4	606	10	Q9XJ29	Q9XJ29	oryza sati
27	187.5	9.3	669	2	Q56151	Q56151	streptomyc
28	187.5	9.3	979	5	Q9Y1K0	Q9Y1K0	plasmodium
29	181.5	9.0	577	10	Q9XJ30	Q9XJ30	oryza sati
30	177.5	8.8	917	10	Q9ZV26	Q9ZV26	arabidopsi
31	171	8.5	351	2	Q9RMT7	Q9RMT7	streptomyc
32	167	8.3	518	2	Q9R5Z6	Q9R5Z6	corynebacte
33	161.5	8.0	718	3	Q94277	Q94277	schizosacch
34	153	7.6	567	2	Q9ZME2	Q9ZME2	helicobacte
35	151	7.5	328	2	Q57527	Q57527	haemophilu
36	150	7.5	428	2	Q05591	Q05591	mycobacteri
37	149	7.4	559	2	Q25066	Q25066	helicobacte
38	132	6.6	348	2	Q67217	Q67217	aquiferi
39	131.5	6.5	336	2	Q46681	Q46681	eschericchia
40	131.5	6.5	336	2	Q46682	Q46682	eschericchia
41	131	6.5	336	2	Q57139	Q57139	eschericchia
42	131	6.5	336	2	Q57446	Q57446	eschericchia
43	129	6.4	336	2	Q46683	Q46683	eschericchia
44	124	6.2	623	2	Q30589	Q30589	streptomyc
45	113	5.6	824	3	Q9Y7F1	Q9Y7F1	aspergillus

## ALIGNMENTS

RESULT	1
09UIA3	
ID	PRELIMINARY;
09UIA3	PRT;
AC	398 AA.
DT	01-MAY-2000 (TIREMBLrel. 13, Created)
DT	01-MAY-2000 (TIREMBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TIREMBLrel. 14, Last annotation update)
DE	NEUTRAL SPHINGOMYELINASE.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=KIDNEY;
RA	Chatterjee S., Han H., Rollins S., Cleveland T.;
RT	"Molecular cloning, characterization, and expression of a novel human
RT	neutral sphingomyelinase.";
RL	J. Biol. Chem. 274:37407-37412(1999).
DR	EMBL: AF069740; AAF19052.1; -.
DR	INTERPRO: IPR000350; -.
DR	PFAM: PF00435; Chotismate_bind; 1.
DR	PRINTS: PR00095; ANTSINHAISEI.
QO	SEQUENCE 398 AA; 43590 MW; 0248D89974C77BC1 CRC64;

Query Match	99.58;	Score 1998.5;	DB 4;	Length 398;
Best Local Similarity	99.78;	Pred. No. 1.6e-146;		
Matches 397;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1

QY	1	MMTHETRALAQSLOOLYALLETTEEGAFATPADDTLFFGICALITATTAALOGA - V	59
Db	1	MMTHETRALAQSLOOLYALLETTEEGAFATPADDTLFFGICALITATTAALOGAVV	60
QY	60	FGAOSFDEOEYPOSELACGFMFPEVWVTTAAADKTFGSDTSVDSFTTWMLOFPVKPQNTV	119
Db	61	FGAOSFDEOEYPOSELACGFMFPEVWVTTAAADKTFGSDTSVDSFTTWMLOFPVKPQNTV	120
QY	120	TTSHVTEVDWMIETETENLIDTLAIDQDTLAKVVFGRQOITLQDSTLRLLAQITIRALAEQANT	179
Db	121	TTSHVTEVDWMIETETENLIDTLAIDQDTLAKVVFGRQOITLQDSTLRLLAQITIRALAEQANT	180
QY	180	YHVVLKRHDELFISSATEPERLVAMSGGOIATAVAVAGTSRRGTGDCADDIALGEALLASQKNR	239

```
|||||
Db 181 YHVLKRDELFI SATPERLVAMSGQIATAVAGTSRRGTGADIALGAEALLASOKNR 240
Qy 240 IEHOYVVASITTRLODVTTSLKVPAMPISLKKKQVOHLTPTTGIOAHLSTAIYDRH 299
Db 241 IEHOYVVASITTRLODVTTSLKVPAMPISLKKKQVOHLTPTTGIOAHLSTAIYDRH 300
Qy 300 PTPALGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYNOTORRAT 359
Db 301 PTPALGVPREALYYIATHEKTPRGLFAGPIGYFTADNSGEFVVGIRSMYNOTORRAT 360
Qy 360 LFGAGIYADSDAOOEYEETGLKFEPMROLKDYNHVE 397
Db 361 LFGAGIYADSDAOOEYEETGLKFEPMROLKDYNHVE 398

RESULT 2
P74053 PRELIMINARY; PRT; 474 AA.
ID P74053
AC P74053;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISCHORI SMATE SYNTHASE.
GN ENTC.
OS Synechocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tanaka S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18129.1; -.
DR INTERPRO; IPR000350; -.
DR PFM; PF00425; choris mate_bind; 1.
DR PRODOM; PD000779; -.
SQ SEQUENCE 474 AA; 52667 MM; 18F8BDB9E5C492C CRC64;

Query Match 19.1%; Score 383; DB 2; Length 474;
Best Local Similarity 31.5%; Pred. No. 1.2e-21;
Matches 112; Conservative 66; Mismatches 150; Indels 28; Gaps 9;

Qy 59 VEGASQFDEOEY--POSELMAGFWFVPEVMYT-----IAADKTFPGSDTVSDFTTW 107
Db 123 IFCGSFDRDRSNRNPFPNSFLFLPQIQVYVKTSGHCLLSWQSLDGNVTVDLVDFIGM 182
Qy 108 LAOFVYKOP--NTVTTSHTVD-----EVDWIERT-ENLIDTLAIDOTLAKVVGROOT 157
Db 183 LSAIRRAQPARDTHTPPSVVAKPRLTGTGEVAKLSKAISLEIHA-QQRLSKVVLATAID 241
Qy 158 LQLSTLTLAQITIRALAEQANTYHYVL--KRDELFI SATPERLVAMSGQIATAVAGT 215
Db 242 LDYGSRLVAVHCLQRLROOYGDICYLFSWNGOGDFCVGASPERLSLHNOCLVTDALAGS 301
Qy 216 SRRGTGADADIALGEALLASOKNRIEHOYVVASITTRLODVTTSLKVPAMPISLKNKOY 275
Db 302 APRDVOVGDRQLGQELHLNPKELREHOAVLDYLLQRLALAGLSQASSL-KLKLKLANIQ 360
Qy 276 HLXPTPTGDIAHLSTVAIVDLRHLPTPALGVPREALYYIATHEKTPRGLFAGPIGYFT 335
Db 361 HLMTQIQARLPRLPHIRLALVQOLHPTPAVAGVPAIAEDLIRRHETFPRLNLAAPLGMWD 420
Qy 336 ADNSGEFVVGIRSMYNOTORRATLFGAGIYADSDAOOEYEETGLKFEPM-ROL 390
Db 421 SEGNAEFTIVGIRSL--SRNRLRYAGAGIYAGSDPLKEVAEIELKIQTLWRSIL 474
```

```
RESULT 3
Q92PCO PRELIMINARY; PRT; 580 AA.
ID Q92PCO
AC Q92PCO;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISCHORI SMATE SYNTHASE (EC 5.4.99.6).
OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Asteridae; eusterids I; Gentianales;
OC Apocynaceae; Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G. DON (MADAGASCAR PERIWINKLE);
RA van Tegelen L.J.P., Moreno P.R.H., Croes A.F., Verpoorte R.,
RA Willems G.J.;
RT "Purification and characterization of isochorismate synthase from
RT elicited cell cultures of Catharanthus roseus (L.) G. Don.";
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ006065; CA006837.1; -.
DR INTERPRO; IPR000350; -.
DR PFM; PF00425; choris mate_bind; 1.
DR PRODOM; PD000779; -.
IS Isomerase.
SQ SEQUENCE 580 AA; 64072 MM; 83F9633BA067312D CRC64;

Query Match 17.9%; Score 360; DB 10; Length 580;
Best Local Similarity 29.9%; Pred. No. 9.4e-20;
Matches 108; Conservative 65; Mismatches 156; Indels 32; Gaps 11;

Qy 60 FGAQSFDEOEY--POSELMAGFWF-VPEV-----MTYIAD--KITFGS---- 98
Db 216 YCAIRFDARPHALPEKAKFGSFYFVQVEPDELHSSMIAATVAMDNALSLYQOAIYA 275
Qy 99 -----DVSDFPTWLAOFKQPNVTYTSHVDEVDWIMRTENLIDTLAIDQT-LAKVVF 152
Db 276 LOTTMEQVSSYVSKLRQDY-SHTSLVSKANIPDRTSMDTLTKRNVLEIGNKSPITKVL 334
Qy 153 GRQ-OTLQSDPTLRALQITRALAEQANTYHYVVKRND-ELFTSATPERLVAMSGQIATA 210
Db 335 ARRSQVITTSIDIDPLAWLSFRKADKDAVOFCLQPREAPAFIGNPDEQFGDQLVFSE 394
Qy 211 AVAGTSRRGTGADADIALGAEALLASOKNRIEHOYVVASITTRLODVTTSLKVPAMPISL 270
Db 395 ALAATRRARESSLDLQMAHDLFSSPKDNHERAIVRENIROKLDICTSVETEPMKSVRK 454
Qy 271 NKQOVHLVPTIGDIAHLSTVAIVDRHLPTPALGVPREALYYIATHEKTPRGLFAGP 330
Db 455 LKRIQHLTARFGRRLSEDEDEKILSLHPTPAVCGFPEDARKKFLAENEMDRGLIYAGP 514
Qy 331 IGYFTADNSGEFVVGIRSMYNOTORRATLFGAGIYADSDAOOEYEETGLKFEPMROL 390
Db 515 VGFPGAGS-DESVGIRSLAIK-DAGALITYAGLVESDPALENGELKASQFMKLM 572
Qy 391 K 391
Db 573 K 573
```

```
RESULT 4
O05851 PRELIMINARY; PRT; 372 AA.
ID O05851
AC O05851;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ENTC.
GN MYC OR RV3215 OR MYCY07D11.11C.
OS Mycobacterium tuberculosis
```

```

Query Match          17.1%; Score 344; DB 2; Length 395;
Best Local Similarity 28.6%; Pred. No. 9,2e-19;
Matches 113; Conservative 65; Mismatches 153; Indels 64; Gaps 12.

QY 19 VALETTEFGAFAFPA-----DTLRFGGAIAATAQAALOGAVF-GAOSFDEQEXP 71
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 31 YRSFTTSGCFARFDEPAVNGDSPSPFOOKLALFADAKAQCIKNPVMGALPFDPHQ-P 89
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 72 QSELMAGFVFVEVWNTIADKIFFGSDIVSDFTTLMAGFVFKQPTVYTSHTDEVDMI 131
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 90 SS-----LYIE-----SW--QSFSSROKQASRRFTR----- 115
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 132 ERTENLIDILAI--DQTLAKVVFGRQOTIOLS---DPLRLAQIIRLAEQANTYHVYLR- 186
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 116 SCSLWVVEGQALPEQTFEQMAARAALATATQYDKVYLSKRLDITTDALIDSGVLEPL 175
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 187 -----HDEL-----FISATPERLVAMSGGQATAAVAGTSRRGTGADDIALGEA 231
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 176 IAQNPVSYNFHPVLDAGGVLLGASPELLLRKDGERRSSIPLAGSARQDEVIDREAGNR 235
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 ILAASKNRIRHXYVASTTRLODYTTSLKAVPAMSLSLKNQOYHLYPTITDIAHLSTV 291
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 236 LASERDRHEHLVLYQAMKEVLRSSSELVAPESSPOLITPLIMLAPPFEGKANSQENA 295
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 292 TAIYVRLHPTPALGCVPREALYYIATHEKTRPGRGLIGYETAQDNGSEFVGTIRSMYV 351
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 296 LTLACLHPTPLSGFPFHQAQVIALEPFPRELFGLGVGWCDSGNGEMVYVITRCAKL 355
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 NQTORATLFPAGAGIVADSDAQOEYEERTGLKEPEM 386
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 356 RENQVR--LEAGAGIVPASSPLGEMRETGVKLS TM 388
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
081522 PRELIMINARY; PRT; 503 AA.
AC 081522
DT 01-NOV-1998 (TREMBLrel. 08. Created)
DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14. Last annotation update)
DE ISOCHORISMATE SYNTHASE.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsids.
RN [1]
RP SEQUENCE FROM N.A.
RA Meng H., Peter G., Pullman G.;
RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF078080; AAC97926.1; -.
DR INTERPRO; IPR000350; -.
DR PFM; PF00425; chorisimate_bind; 1.
DR PRODOM; PD000779; -.
SQ SEQUENCE 503 AA; 55359 MW; 50FAFA42DE862742 CRC64;

Query Match          15.3%; Score 307.5; DB 10; Length 503;
Best Local Similarity 28.2%; Pred. No. 8,6e-16;
Matches 106; Conservative 59; Mismatches 158; Indels 53; Gaps 11.

QY 21 ALETFEFGA-YRATPADTLRFGAIAATAKTA-----QALQAVFGAQSFDEQETPOS 73
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 147 AVEWMPFGAFYFSPQVFEFNEGSSMLAATITAMDELSWTLENAIEALQ----- 196
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 ELMAGFVFWPEVWNTIADKIFFGSDIVSDFTTLMAGFVFKQPNVYTSHTDEVDMIER 133
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 197 -----ETMLQSSVYMKLNRNRSLG-----VSVLSKNHVPFKGAVRPPA 233
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 TENLIDTL-AIDQTLAKVVFGRQOTIOLSDT-----LRLAQIIRLAEQANTYHVYLRH 187
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
Db 234 VEKALEMINOKSSPLNKKVVLARNRSRI-TTDDTDIDPIAWLAQOLR---EGHDVAYOFCLOPP 289
OY 188 D-ELFISATPERLVAMSGGOIATAAAGTSRRGTGADGADIDALGEALLASOKNRIEHOYV 246
    || ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 GAFAPIGNTERLRFQRLQGVCSSEALATAPRAASSARDEIEIDLTLSKRDLESIYR 349
OY 247 ASITTRLODVTTSLKYPAMPISLKNKOVOHLVPTITGDIAHLVTAIVDRLHPPTALGG 306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 ENIREKLNIGICDRVVVYKPOKTVKRLARVOHLVLSQLAGRLTKEDDEXKILALHPPTAVCG 409
OY 307 VPREALIYATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORALLFPAGAGI 366
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 LPAEERALLIKETESFDRGMVAGPIGFFGGEES-EFVAGIRSAVY-EKGIGALIVAGTGI 467
OY 367 VADSDAQOYEYETGLK 382
    || || : : : : :
Db 468 VAGSDPSSSEWNEIDLK 483

RESULT 7
ID 09S7H8 PRELIMINARY: PRT: 503 AA.
AC 09S7H8:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISOCHORISMATE SYNTHASE (ICSI).
GN FIM20.39 OR F25A4.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Ienz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Altafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Hong B., Huizer L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F25A4 sequence.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC011765; AAF1594.1; -.
DR EMBL: AC008263; AAD55272.1; -.
DR INTERPRO: IPR000350; -.
DR PFMK: PF00425; Chorismate_bind; 1.
DR PRODOM: PD000779; -.
SQ SEQUENCE 503 AA: 55268 MW: 8CFC36477D54F61E CRC64;

Query Match 15.3%; Score 307.5; DB 10; Length 503;
Best Local Similarity 28.2%; Pred. No. 8; 6e-16;
Matches 106; Conservative 59; Mismatches 158; Indels 53; Gaps 11;

OY 21 ALETFEFGA-YEATPADTLRFPGIAGIATAKTA-----QALQAGVFGAOSFDEQYPPQS 73
    : : ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 AVEWPFQAFYFSVPOVENEGSSMLAATIAMDELSTWLENALEAQ----- 196
OY 74 ELMAGFVWPEVMTIADKITFGSDTVVSDFTTWLAQFVPKOPNTVTTSHTDEVDMIER 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 -----ETMLQVSSVVMKLRNRSIG-----VSYSKNNHVTPTKGYFPA 233
OY 134 TENLIDTL-AIDQTLAKYVFGROQLQLOSDT-----LRLAQITRLALAEANITYHVLKHN 187
    : : ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 VEKALEMINOKSSPLNKKVVLARNRSRI-TTDDTDIDPIAWLAQOLR---EGHDVAYOFCLOPP 289
```

```
OY 188 D-ELFISATPERLVAMSGGOIATAAAGTSRRGTGADGADIDALGEALLASOKNRIEHOYV 246
    || ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 GAFAPIGNTERLRFQRLQGVCSSEALATAPRAASSARDEIEIDLTLSKRDLESIYR 349
OY 247 ASITTRLODVTTSLKYPAMPISLKNKOVOHLVPTITGDIAHLVTAIVDRLHPPTALGG 306
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 ENIREKLNIGICDRVVVYKPOKTVKRLARVOHLVLSQLAGRLTKEDDEXKILALHPPTAVCG 409
OY 307 VPREALIYATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORALLFPAGAGI 366
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 LPAEERALLIKETESFDRGMVAGPIGFFGGEES-EFVAGIRSAVY-EKGIGALIVAGTGI 467
OY 367 VADSDAQOYEYETGLK 382
    || || : : : : :
Db 468 VAGSDPSSSEWNEIDLK 483

RESULT 8
ID P95475 PRELIMINARY: PRT: 391 AA.
AC P95475:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ISOCHORISMATE SYNTHASE.
GN FBSC.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS374;
RA Mercado-Blanco J., Olsson P.E., van Loon L.C., Bakker P.A.H.M.;
RT "Analysis of a gene cluster involved in salicylate biosynthesis and
RT detection of a novel siderophore in Pseudomonas fluorescens WCS374.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y09356; CAA70528.1; -.
DR INTERPRO: IPR000350; -.
DR PFMK: PF00425; Chorismate_bind; 1.
DR PRODOM: PD000779; -.
SQ SEQUENCE 391 AA: 43102 MW: 6654B0901C7D56A1 CRC64;

Query Match 15.2%; Score 304.5; DB 2; Length 391;
Best Local Similarity 28.6%; Pred. No. 1e-15;
Matches 108; Conservative 46; Mismatches 160; Indels 63; Gaps 10;

OY 68 QEYPOSELMAGFVFP---EVMVTIADKITE-----GSDTVSDFTTWLAQ----- 110
    : : ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 EEVQLAEVGRSFSFTSGDRELAVTGMLORIEFPALIGDDANSLFQQTINQALDRAREAQ 71
OY 111 -----FVFKOPNTVTTSHTDEVDMIER-----TENLIDTLAID 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 SNPIIVGAIPIFPAPBSCL---YIPEHAQWRTDIHAKTGMSTPLPELIEQKNIPDEQARK 128
OY 145 OTLAKYVFG-----RQQLQLSDTLRLAQIIRALDAQNT-----YHVLKRRD-EL 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 RAVEHAIVVFRHSDYKKAIVSQRRELIFANDVDVSLQHNLLKQNPDSGTHFRVPMPDGTT 188
OY 191 FTSATPERLVAMSGGOIATAAAGTSRRGTGADGADIDALGEALLASOKNRIEHOYVASTT 250
    : : ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 LIGVSPPELLVRRREGSLSSNPPLAGSKAKRNADPADRNMDWLTSKDIYEHQFYQDIV 248
OY 251 TRLODVTTSLKYPAMPISLKNKOVOHLVPTITGDIA-AHLVTAIVDRLHPPTALGGVR 309
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 SOLGKICTOLNVPORSLSTPALMHLSTRIGETLADPAVSALQLCRLHPPTAVCGPPT 308
OY 310 EALYIYATHEKTPRGIFAGPIGFTADNGSEFVVGIRSMYVNOTORALLFPAGAGI 369
    : : ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 ERARRLIRFVEPERGLFTGMVGMCDAGNGEVMYVTRKGVRRNKR--LFGAGIVFA 366
OY 370 SDAQOYEYETGLKFEPM 386
```



Db 367 SSPDSEMAEVOYTKLGT 383

RESULT 9

Q97M09 PRELIMINARY; PRT; 449 AA.

AC Q97M09: 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE HYPOTHEtical 52.3 KDA PROTEIN.

GN MENF.

OS Cyanidium caldarium (Galdieria sulphuraria).

OC Chloroplast.

CC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;

CC Cyanidium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-RK1:

RA Gloeckner G., Rosenthal A., Valentin K.;

RT "Reconstruction of the ancient red algal plastid genome: Structure, gene repertoire, and synteny."

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF022186; AAF12987.1; -

DR INTERPRO: IPR000350; -

DR PFAM: PF00425; choriSmate\_bind; 1.

DR Hypothetical protein; Chloroplast.

KW SEQUENCE 449 AA; 52316 MW; B7771DE473ED2BF3 CRC64;

SO

Query Match 14.7%; Score 295.5; DB 8; Length 449;

Best Local Similarity 31.5%; Pred. No. 6.2e-15;

Matches 78; Conservative 49; Mismatches 116; Indels 5; Gaps 4;

QY 146 TLAKVFGROQTQLGSDTLRLAQLIRALA-EQANTYHVYKHHDEL-FISATPERIVAMS 203

DB 204 SLEKIVLARRKLEFETDLSLITQOLKHKQYNSYNGRLFTNTNGFLAFSPERIKYRK 263

QY 204 GGQIATAAVAGTSRRGTGADGADIALGALLASOKNRIEHQYVASTITRLQDVTSLKVP 263

DB 264 TNLITSEAVAGTRLRGRQKHLRLAFLNLICSKDFENIYKQVLVNTKNGINIVKQ 323

QY 264 AMPSLKKQYOHLYTPITGDIAHLVTAIVDRHLHPALGCVPREALYYIAHTEKTP 323

DB 324 LHYSTMOTSNVOHLISRLQAKLSNMSDQKILRNLHPPALIGSPQQAASKCIPLEFNRY 383

QY 324 RGLFAGPIGYFTADNSGEFVVGIRSMYVNOTQRATLTFAGAGIVADSDAQOYEETGLKF 383

DB 384 RCTVAPSPYGFYTDNS-SMKISIRSAFIN-KRKILIIAGCGILNSIYLKEMNEVENKM 440

QY 384 EPMROLK 391

DB 441 NNFTIIE 448

RESULT 10

Q9VIG5 PRELIMINARY; PRT; 434 AA.

AC Q9VIG5: 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE ANTHRANILATE SYNTHASE COMPONENT I (TRPE).

GN PAB2045.

OS Pyrococcus abyssi.

CC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ORSAY;

RA Heilig R.;

RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ248284; CAB49384.1; -

DR INTERPRO: IPR000350; -

DR PFAM: PF00425; choriSmate\_bind; 1.

DR PRINTS: PR00095; ANTSNTHASEI.

SO SEQUENCE 434 AA; 49058 MW; 1FFBF19D48BD193C CRC64;

Query Match 12.5%; Score 251.5; DB 1; Length 434;

Best Local Similarity 27.4%; Pred. No. 1.5e-11;

Matches 69; Conservative 59; Mismatches 113; Indels 11; Gaps 5;

QY 149 KVAFGRQOTQLGS-DLRLAQLIRALA-EQANTYHVYKHHDEL-FISATPERIVAMSGGQI 207

DB 191 QVVLSEYVLRSDVDPELYKRLISINSPYTFLEFK--LVGASPEPTMGSGVGRV 247

QY 208 ATAAVAGTSRRGTGADGADIALGALLASOKNRIEHQYVASTITRLQDVT--SLKVPAM 265

DB 248 RINPIAGTAPRGKTPREDEIRRLRLSDKERAEHMLVDLANDYKSKSPGVLYNRP 307

QY 266 PSLKKQYOHLYTPITGDIAHLVTAIVDRHLHPALGCVPREALYYIAHTEKTPRG 325

DB 308 FVIKYSHVOHIESEVIGELADQKDFDAIEASFPAFTLGAPKIRAMEIIDLEKSRRR 367

QY 326 LFRAGPIGYFTADNSGEFVVGIRSMYVNOTQRATLTFAGAGIVADSDAQOYEETGLKFP 385

DB 368 VYGAIGYFSTIGYADFAIRMAELE--GRNAHVAGAGIVADSIPEKEFYETENK--- 422

QY 386 MRQLKDYNHVE 397

DB 423 MKAVLKAFGVITQ 434

RESULT 11

Q9S2U2 PRELIMINARY; PRT; 502 AA.

AC Q9S2U2: 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE PUTATIVE ANTHRANILATE SYNTHASE COMPONENT I.

GN SC46.12C.

OS Streptomyces coelicolor.

CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Redenbach M., Kiese H.M., Denapalte D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL MOL. Microbiol. 21:77-96(1996).

DR EMBL: AL096884; CAB51435.1; -

DR INTERPRO: IPR000276; -

DR INTERPRO: IPR000350; -

DR PFAM: PF00425; choriSmate\_bind; 1.

DR PRINTS: PR00095; ANTSNTHASEI.

DR PROSITE: PS00237; G\_PROTEIN\_RECEPTOR; UNKNOWN\_1.

SO SEQUENCE 502 AA; 54164 MW; 0BCD21DDE8983436 CRC64;

Query Match 11.3%; Score 227; DB 2; Length 502;

Best Local Similarity 26.3%; Pred. No. 1.4e-09;  
Matches 74; Conservative 50; Mismatches 141; Indels 16; Gaps 5;

QY	108	LAGEVAPROPNVTVSHTYDEV-----DWIERENLIDTIAIDLOTLAKVVFGEQOOLQLS	161
Db	208	LTRAVACPPALPSELPEETATLMGCGFDPODAVEDIERIRAGAEVGVYSQREFTPCA	267
QY	162	DTLRILAQTIRALAAEQANTYHVHLKRHDELFI SATPERLVAMSGQIATAVAAGTSRGTD	221
Db	268	SALDVRVRLRTANSPMYLLRI LDGFC- -VGSSPEALTVEEDGRAMVHPDIAGRPGAT	325
QY	222	GADDIALGCEALLAQSKRRIEHOVVVASITRRLDVYT--SLKYPAMSLLKNQVOHLT	279
Db	326	PREDOSLADELALPAKKEAEHLMLVYDLGRNDLCVGCPEGSEVADVSVERYSHVMITYS	385
QY	280	PITGDIAHALSVTAIVDRLRHTPALGCVPREAALYYIATHEKTPRGFLFAPIGYFTADNS	339
Db	386	TVTGRVAVAGRTAFDVLTWACFPAGCTLSAPRPRAQLIIDLEPSRGLGYGCVCYLDPAGD	445
QY	340	GEFVVGIRSMKVNOTORRALTF--AGAGIYAADSQAQOEYE E	378
Db	446	SDTAIAIRITALL---RDGAIVYVAGGVAAVDSPVAEDETE	482

RESULT	12
P72539	
ID	P72539
AC	P72539
DT	01-FEB-1997 (TREMBLrel. 02, Created)
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	PAPA.
GN	PAPA.
OS	Streptomyces pristinaespiralis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE: P7197164.
RA	Blanc V., Gil P., Bamas-Jacques N., Lorenzon S., Zagorec M.,
RA	Schleuinger J., Bisch D., Blanche F., Debussche L., Crouzet J.,
RA	Thibaut D.;
RT	"Identification and analysis of genes from Streptomyces
RT	pristinaespiralis encoding enzymes involved in the biosynthesis of the
RT	4-dimethylamino-L-phenylalanine precursor of pristinamycin I.";
RL	Mol. Microbiol. 23:191-202(1997).
DR	EMBL: U60417; AAC44866.1; -.
DR	INTERPRO: IPR000350; -.
DR	INTERPRO: IPR000991; -.
DR	INTERPRO: IPR002385; -.
DR	PFAM: PF00425; chorismate_bind; 1.
DR	PFAM: PF00117; GATase; 1.
DR	PRINTS: PRO0095; ANTSNTHASE1.
DR	PRINTS: PRO0096; GATASE.
DR	PRINTS: PRO0097; ANTSNTHASE1.
DR	PROSITE: PS00442; GATASE_TYFE_I; 1.
DR	PRODOM: PD00779; -; 1.
QO	SEQUENCE 719 AA: 77363 MW: 85850000.000000 BEEBFCBDA91D8368 CRC64:

Query Match	10.9%;	Score 219;	DB 2;	Length 719;
Best Local Similarity	25.7%;	Pred. No. 1e-08;		
Matches 77;	Conservative 53;	Mismatches 142;	Indels 28;	Gaps 7;

[illegible][illegible]

RESULT	13		
094582			
ID	094582	PRELIMINARY;	PRT; 489 AA.
AC	094582;		
DT	01-MAY-1999	(TREMBLrel. 10, Created)	
DT	01-MAY-1999	(TREMBLrel. 10, last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, last annotation update)	
DE	ANTHRANILATE SYNTHASE COMPONENT I.		
GN	SPCC1442.09.		
OS	Schizosaccharomyces pombe (fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;		
OC	Schizosaccharomycetaceae; Schizosaccharomyces.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-972H-;		
RA	Lyme M., Rajadream M.A., Bartell B.G., Volckaert G.;		
RL	Submitted (OCR-1998) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AL031966; CAA21443.1; -		
DR	INTERPRO: IPR00350; -; -		
DR	PFAM: PF00425; chorismate_bind. 1.		
DR	PRINTS: PR00095; ANTSYNTHASE1.		
DR	PRODOM: PD000779; -; 1.		
SQ	SEQUENCE 489 AA; 54959 MW; 442D552BA59EBA31 CRC64;		

Query Match	10.9%;	Score 218.5;	DB 3;	Length 489;
Best Local Similarity	33.7%;	Pred. No. 6.2e-09;		
Matches	65;	Conservative 35;	Mismatches 84;	Indels 9;
				Gaps 5

```

QY 192 ISNPERLYAMNSGQIAIAAAVAGSRCTDDADDILAGENLLASOKNRIEQYV---AS 248
    | : | | | | : | : | | | : | : | | | | : | : |
Db 275 IGASELLVYKSEBHGRIINHPIAGTVPRCKTKEEDEAAVAKDILLASVKDRAEYHVLIDLAN 334
    | : | : | : | : | : | : | : | : | : | : | : |
QY 249 IITTRQDVTYTSKLYPAMPSSLKNNQVOHLYPIPGDIAAHLSTVAIYDRLPHPTALGCV 308
    | : | : | : | : | : | : | : | : | : | : | : |
Db 335 DYSKRCDLDTI-SYDKLMTTIEKFSHYQHLYSQVSGVLRPKDTRDPARSRIFPACTVSGSP 393
    | : | : | : | : | : | : | : | : | : | : | : |
QY 309 REALYIATHEKTPRGLEAPIG-YFTDANSGEFVVGIRSM-VYNQTRRATLEAGAGI 366
    | : | : | | | : | : | : | : | : | : | : | : |
Db 394 KVRATQIYVYGLEKEKRGTYAGAVGWMGEDNDMDTCAIRTMVYKDGK---VYLQAGGI 450
    | : | : | : | : | : | : | : | : | : | : | : |
QY 367 VADSPAQOEYEET 379
    | | | | | | |
Db 451 VFDSDQDEDEYET 463
    | | | | | | |

```

RESULT	14
09X918	
ID	09X918
AC	PRELIMINARY;
DT	09X918;
DT	01-NOV-1999 (TREMBlrel. 12, Created)
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT	01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE	SALICYLATE SYNTHETASE, IRP9.
GN	IRP9.
OS	Yersinia enterocolitica.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Yersinia.
CC	[1]
RN	SEQUENCE FROM N.A.
RP	
RA	Ratkin A.V., Pelludat C., Noetting C., Schubert S., Jacobi C.,



---

**THIS PAGE BLANK (USPTO)**